Biochemical Characterization of the CHD1, CHD5 and SMARCAD1 Chromatin Remodellers

Michael Lim

University College London

and

Cancer Research UK London Research Institute

PhD Supervisor: Dr Jesper Q. Svejstrup

A thesis submitted for the degree of

Doctor of Philosophy

University College London

March 2016

Declaration

I, Michael Lim, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

By altering nucleosome positioning and composition, chromatin remodellers can modulate genomic accessibility, thereby regulating DNA metabolic processes. However, many remain poorly characterized, especially regarding the mechanisms underlying their regulation and genomic specificity.

Some chromatin remodellers form multi-subunit complexes that change in subunit composition during development. It was reasoned that identifying tissue-specific subunits of chromatin remodellers would help elucidate their regulation. Thus, endogenous CHD1 and CHD5 were purified from mouse embryonic stem cells and brains respectively. However, detailed mass spectrometry analysis failed to reveal any novel tissue-specific complex subunits of either enzyme.

The poorly characterized chromatin remodeller, SMARCAD1, is implicated in the maintenance of heterochromatin and double-stranded DNA break repair. SMARCAD1 constitutively forms a stoichiometric interaction with KAP1, a transcriptional co-repressor. The SMARCAD1-KAP1 complex was reconstituted *in vitro* with pure recombinant proteins. Although both SMARCAD1 and the SMARCAD1-KAP1 complex display nucleosome-stimulated ATPase activity, SMARCAD1 exhibits greater activity than the complex, suggesting that KAP1 is a negative regulator.

Beyond a conserved SWI2/SNF2-like ATPase domain, SMARCAD1 features tandem CUE domains, which in other proteins, function as ubiquitin-binding domains. In cells, we find that point mutations affecting the SMARCAD1 CUE domains abrogate the SMARCAD1-KAP1 interaction. Significantly, using recombinant proteins that are not ubiquitylated, the SMARCAD1-KAP1 interaction was shown to be a direct, CUE domain-dependent, protein-protein interaction, involving the first SMARCAD1 CUE domain and the RBCC domain of KAP1. Unexpectedly, *in silico* approaches fail to identify a ubiquitin-homology domain within KAP1, suggesting either that KAP1 possesses an atypical ubiquitin-homology domain or that the first SMARCAD1 CUE domain uniquely recognizes a

ligand structurally distinct from ubiquitin. An attempt to determine the co-crystal structure of the SMARCAD1-KAP1 interaction interface is described. Overall, the biochemical and structural characterization of the SMARCAD1-KAP1 complex offer mechanistic insight into the regulation of a chromatin remodeller with important roles in DNA metabolism.

(300 words)

Acknowledgement

I am immensely grateful to my PhD supervisor, Jesper Svejstrup, for the opportunity to work in his lab, while being afforded prodigious freedom to dictate the nature and direction of my projects. In addition to scientific guidance, I have benefitted greatly from his eternal optimism and unwavering support.

The Svejstrup Lab has been an extremely conducive and accommodating environment in which to work, for which I would like to thank all the members of the lab. In particular, however, I need to thank Andy Ehrensberger for being an incredible mentor, especially for the generosity with which he offered me his time and expertise, not to mention nurturing in me a persisting appreciation for reductionistic science. I would also like to thank Hannah Williams, with whom I have collaborated on the SMARCAD1-KAP1 project, for her continuing advice and assistance. I would also like to acknowledge my thesis committee – Steve West and Simon Boulton – for useful scientific discussions. My collaborator, Matthieu Gérard, has also been generous in providing reagents and materials.

I would like to thank Peter Cherepanov and the entire Cherepanov Lab for welcoming me into their midst to attempt X-ray crystallography. Specifically, I would like to thank Valerie Pye for guiding me through the crystallization process, and Dan Maskell for giving me copious quantities of nucleosomes. The scientific infrastructure at the London Research Institute has also greatly facilitated this study – most notably, the Protein Analysis & Proteomics, Cell Services & Media, Fermentation and Peptide Chemistry core facilities. I would also like to acknowledge the financial support provided by Cancer Research UK and UCL.

I am also grateful to the UCL MBPhD programme for offering me the opportunity to undertake a PhD at this stage of my medical training. Gordon Stewart, who embodies the programme, has been especially supportive.

Finally, I owe a deep debt of gratitude to my family and friends for their enduring support and encouragement.

Table of Contents

Table of Figures

List of Tables

Abbreviations

ACF	ATP-utilizing chromatin assembly and remodelling factor
β-ME	β-mercaptoethanol
Bp	base pairs
CHD (1-9)	Chromodomain helicase DNA-binding protein (1-9)
ChIP-chip	Chromatin immunoprecipitation-microarray chip
ChIP-seq	Chromatin immunoprecipitation-DNA sequencing
Cryo-EM	Cryogenic electron microscopy
CUE domain	Coupling of ubiquitin conjugation to endoplasmic reticulum degradation domain
CV	Column volume
dsDNA	Double-stranded DNA
DOC1	Deleted in oral cancer 1
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DSB	Double-strand breaks
DTT	Dithiothreitol
ESC	Embryonic stem cell
FRET	Fluorescence resonance energy transfer
GST	Glutathione S-transferase
H2Aub	Mono-ubiquitylation of histone H2A; a repressive mark
H2Bub	Mono-ubiquitylation of histone H2B; an active mark
H3K4me3	Histone H3 tri-methylated at lysine 4; an active mark
H3K9me3	Histone H3 tri-methylated at lysine 9; a repressive mark
H3K27me3	Histone H3 tri-methylated at lysine 27; a repressive mark
H4K16ac	Histone H4 acetylated at lysine 16
HDAC1	Histone deacetylase1
HP1	Heterochromatin protein 1
HSA domain	Helicase-SANT-associated domain
HSS domain	HAND-SANT-SLIDE domain
iBAQ	Intensity-based label-free absolute-quantification
IP	Immunoprecipitation

IPTG	Isopropyl β-D-1-thiogalactopyranoside
KAP1	KRAB-associated protein 1
Kb	Kilo base pairs
K_d	Dissociation constant
KRAB domain	Krüppel-associated box domain
LC-MS	Liquid chromatography-mass spectrometry
LIF	Leukaemia inhibitory factor
LRI	London Research Institute
MBD2	Methyl-CpG-binding domain 2
MEF	Mouse embryonic fibroblast
MWCO	Molecular weight cut-off
MTA1	Metastasis-associated 1
NLS	Nuclear localization signal
NMR	Nuclear magnetic resonance
NuRD complex	Nucleosome remodelling and histone deacetylation complex
OD_{600}	Optical density at 600nm
PCR	Polymerase chain reaction
PEP	Posterior error probability
PRC	Polycomb repressive complex
PTM	Post-translational modification
RBBP4	Retinoblastoma binding protein 4
RBCC domain	RING, B-box and coiled-coil domain
RNAPII	RNA polymerase II
RT	Room temperature
SANT domain	Swi3, Ada2, N-CoR and TFIIB domain; a DNA-binding domain
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SHL	Superhelix location
SILAC	Stable isotope labeling by amino acids in culture
SIM	SUMO-interacting motif
Sir	Silencing information regulator
SLIDE	SANT-like ISWI domain; a DNA-binding domain
TSS	Transcription start site
TRIM protein	Tripartite motif-containing protein
TTS	Transcription termination site

Chapter 1. Introduction

The focus of this study was the biochemical characterization of the CHD1, CHD5, and SMARCAD1 chromatin remodellers, with the objective of elucidating upon their regulatory mechanisms. Thus, this introduction will present a general overview of chromatin – as befitting the setting in which chromatin remodellers exert their activity – and chromatin remodellers, before focusing on these three specific enzymes. Nonetheless, in keeping with the biochemical theme of this work, these topics will be approached with an emphasis on structural and *in vitro* studies that offer reductionistic insights into the fundamental properties of chromatin and into the basic mechanisms of chromatin remodelling.

1.1 Chromatin

1.1.1 Structural Organization of Chromatin

Each human nucleus contains nearly two metres of linear DNA (Clapier and Cairns, 2009). In the cell, this genetic material exists in the form of a complex of DNA and proteins, termed chromatin. The fundamental structural subunit of chromatin is the nucleosome, which is composed of 146 base pairs (bp) of DNA wrapped around a histone octamer core – itself comprised of two copies of each of the core histones, H2A, H2B, H3 and H4 – as approximately 1.7 superhelical turns (Figure 1.1) (Kornberg, 1974, Luger et al., 1997). Thus, a length of linear DNA can be wound into an 11nm fibre resembling beads on a string, with each nucleosome being separated by 20-75bp of linker DNA. Yet, by itself, this mode of packaging is insufficient to achieve, for instance, the 10,000-20,000-fold compaction that transforms linear DNA into chromosomes. Hence, a hierarchical organization akin to that of proteins has been proposed for chromatin, with the 11nm fibre being comparable to the primary structure of proteins, and the overall chromatin structure being dependent on higher order packaging (Woodcock and Ghosh, 2010).

Figure 1.1 – Structure of the Nucleosome

A. The structure of the nucleosome core particle illustrates how 146bp of DNA is wound around a histone octamer in approximately 1.7 superhelical turns. The DNA backbone is coloured in brown and turquoise, histone H2A in yellow, H2B in red, H3 in green and H4 in blue. The image on the left is the view of the nucleosome looking down the superhelical axis, while the image on the right is the view perpendicular to the superhelical axis.
B. Half the nucleosome is shown here to illustrate the path of nucleosomal DNA. This is a view down the superhelical axis, and the colour scheme is the same as in **A**. The central base pair is denoted superhelix location (SHL) 0. Each SHL increment represents one complete turn of the DNA double helix from SHL0. The positions SHL 0-7 are labelled.
The images were adapted from (Luger et al., 1997).

Extending this analogy, the 30nm fibre that results from the formation of inter-nucleosomal histone contacts is comparable to the secondary structure of proteins. This conformation – a compacted chromatin fibre 30nm in diameter – is adopted by polynucleosomal arrays *in vitro* when in the presence of buffers containing high concentrations of cations, which enables chromatin to overcome self-repulsion due to the polyanionic nature of DNA. While the 30nm fibre has long been described and is observed reproducibly, controversy abounds about its physiological relevance since it has yet to be directly observed in cells (Luger et al., 2012). Nonetheless, two plausible models have been proposed for the structure of 30nm fibre (Woodcock and Ghosh, 2010, Luger et al., 2012). The first, the "solenoid" or "one start" model, predicts that the 11nm fibre of nucleosome beads is coiled into a helix, whereby the linker DNA continues the helical trajectory of the nucleosomal DNA (Figures 1.2A and B). Thus, the nucleosomes are arranged such that inter-nucleosomal histone contacts occur between consecutive nucleosomes (Robinson et al., 2006, Woodcock and Ghosh, 2010, Luger et al., 2012). The second, the "zigzag" or "two start" model, features nucleosomes arranged in a zigzag pattern, whereby alternating nucleosomes interact with each other (Figures 1.2C and D).

Figure 1.2 – Model of the 30nm Fibre
A. The "one-start" or "solenoid" model is based on inter-nucleosomal histone contacts between consecutive nucleosomes.
B. This schematic representation of the "one-start" model shows the outcome of the intervening linker DNA continuing the helical trajectory of the nucleosomal DNA – namely, the 11nm fibre of nucleosome beads being coiled into a helix.
C. The "two-start" or "zig-zag" model is based on inter-nucleosomal histone contacts between alternating nucleosomes (i.e. N1, N3, N5 etc).
D. This schematic of the "two-start" model emphasizes how the two different models result in chromatin fibres that differ in width and in the trajectory of the linker DNA.
The image was taken from (Luger et al., 2012).

The “two start” model is favoured by structural studies, such as the crystal structure of a reconstituted tetra-nucleosome, albeit one that lacks the H1 linker histone, has short linker DNA, and was obtained under non-physiological crystallization conditions (that included 90mM magnesium chloride) (Schalch et al., 2005). Crucially, however, this mode of structural organization was also observed in the recent cryogenic electron microscopy (cryo-EM) structure at ~11Å resolution of a 30nm fibre reconstituted *in vitro* with recombinant *Xenopus laevis* histone octamers, histone H1, and a DNA fragment containing twelve 601 nucleosome-positioning sequences (Figure 1.3A). The 30nm fibre forms a left-handed double helix, comprised of repeating tetra-nucleosomal structural units, within which nucleosomes zigzag back and forth as a “two start” helix (Figure 1.3B). Thus, the tetra-nucleosomal unit is arranged into two stacks of nucleosomes joined by linear linker DNA. As the linker DNA is straight and extended, the diameter of the 30nm fibre increases as the length of linker DNA increases (Song et al., 2014).

An important determinant of higher order chromatin packaging in metazoans is the H1 linker histone, which interacts with approximately 20bp of linker DNA. Through its lysine-rich carboxy-terminus, H1 binds strongly to DNA. Notably, nucleosomal arrays lacking histone H1 are refractory to the formation of 30nm fibres (Woodcock and Ghosh, 2010). The cryo-EM structure of the 30nm fibre reveals that the linker histone H1 associates with nucleosomes with a one to one stoichiometry, and that it is positioned asymmetrically towards the side of the nucleosome that contacts the next tetra-nucleosomal unit. This allows histone H1 to dictate that each tetra-nucleosomal unit is rotated roughly 50° and translated by approximately 70Å along the stack axis relative to the preceding unit; these positional constraints by H1 govern the spiral twisting of the 30nm fibre into a left-handed superhelix (Figure 1.3C) (Song et al., 2014).

Figure 1.3 – Cryo-EM Structure of the 30nm Chromatin Fibre

A. This image is the cryo-EM reconstruction of the 30nm fibre. Each tetra-nucleosomal structural unit is highlighted in a different colour. In this figure, nucleosomes are labelled N1-N12.

B. This schematic representation shows the straight linker DNA connecting the two stacks of nucleosomes, illustrated as discs.

C. This pseudo-atomic model of the 30nm fibre was built by docking 4 copies of the 12-nucleosome structure, though the structure of histone H1 was not included here. This model illustrates the left-handed superhelix formed by the 30nm fibre. As in **A**, the tetra-nucleosomal structural units are coloured in blue, orange or purple for clarity. The scale bar is 11nm in length

The images were taken from (Song et al., 2014).

Apart from unique exceptions such as metaphase chromosomes, our understanding of higher order chromatin structure beyond the 30nm fibre ventures largely into the realms of speculation. As investigation into this area relies disproportionately upon light microscopy, one main limitation is the inherent resolution limit of approximately 250nm imposed by the properties of light (Woodcock and Ghosh, 2010). Nonetheless, further evidence for non-random higher order chromatin organization is provided by the existence of long-range chromatin interactions that mediate regulatory functions. This phenomenon is arguably best illustrated by enhancers, which are DNA sequences that activate transcription from distant promoters separated by up to a hundred kilo base pairs (kb). This is possible as chromatin looping allows enhancers and their cognate promoters to be in close three-dimensional spatial proximity. Many enhancers are essential for developmental processes, emphasizing that chromatin organization is tightly regulated and represents much more than a means for DNA compaction (Kulaeva et al., 2012, Shlyueva et al., 2014).

Initially a cytological description of areas of interphase chromatin that stain intensely with basic dyes, heterochromatin does not refer *per se* to a type of higher order chromatin structure, but instead describes an area of highly compacted chromatin with limited transcriptional activity that tends to adopt a peripheral location in the nucleus. Correspondingly, euchromatin is loosely compacted and has high transcriptional activity (Heitz, 1928, Woodcock and Ghosh, 2010). Constitutively heterochromatic regions always remain highly compacted and transcriptionally inactive; these tend to be gene-poor, late-replicating portions of the genome that are often marked by tri-methylation of lysine 9 of histone H3 (H3K9me3). Alternatively, facultative heterochromatin may de-condense, allowing transcription in a context-specific manner, for example during specific developmental or cell cycle stages. Facultative heterochromatin is enriched in tri-methylated lysine 27 of histone H3 (H3K27me3), a histone modification deposited

by the Polycomb Repressive Complexes (PRC). In contrast, however, transcriptionally competent euchromatin is characterized by “active” histone marks such as tri-methylation of lysine 4 of histone H3 (H3K4me3) and hyper-acetylation of the tails of histones H3 and H4 (Trojer and Reinberg, 2007).

1.1.2 Modulation of Genomic Accessibility

Packaging of DNA into higher order chromatin structures is intrinsically at odds with genome accessibility. Thus, the complex molecular machines that replicate, transcribe and repair DNA are obstructed from accessing regions of the genome that are compacted into chromatin (Clapier and Cairns, 2009). Even without contending with higher order chromatin structures, nucleosomes represent imposing obstacles for DNA metabolic enzymes. For example, RNA polymerase II (RNAPII) is inhibited from initiating transcription when the promoter sequence is wrapped around a nucleosome (Lorch et al., 1987). It is estimated that recognition sequences within nucleosome-occupied promoter regions are 10-20 fold less accessible to protein factors than identical regions of DNA lacking nucleosomes (Iyer and Struhl, 1995, Struhl and Segal, 2013). Therefore, the packaging DNA into chromatin represents an important level at which DNA metabolic processes may be regulated.

Local modulation of genomic accessibility can be accomplished by dynamic and reversible alterations in chromatin structure mediated principally by two mechanisms – covalent modifications of histone tails and ATP-dependent chromatin remodelling (Clapier and Cairns, 2009, Tessarz and Kouzarides, 2014). Post-translational modifications of specific residues on the exposed tails of core histones may either act directly by altering the stability of nucleosomes, or indirectly by recruiting additional factors (Tessarz and Kouzarides, 2014). In contrast, SWI2/SNF2-like chromatin remodelling ATPases couple the energy released from ATP hydrolysis to repositioning, ejecting or modifying the composition of nucleosomes, for example by incorporating histone variants in place of the canonical core histones (Clapier and Cairns, 2009).

A characteristic feature of active chromatin is histone hyper-acetylation. Histone H3 has been reported to be acetylated at specific lysine residues, including H3K56 that is located near the DNA entry and exit site on nucleosomes, and H3K122 that maps to the nucleosomal dyad axis. As acetylation neutralizes the positive charge of lysine residues, it is eminently conceivable that the histone-DNA interaction is weakened by these modifications (Tessarz and Kouzarides, 2014). While acetylation of H3K56 is probably insufficient to alter the overall nucleosome structure, by weakening histone-DNA contacts at the entry/exit point, it is capable of preventing oligomerization of sub-saturated nucleosomal arrays. Consequently, it has been proposed that acetylation of lysine 56 is a mechanism for the establishment of nucleosome-free regions (Watanabe et al., 2010). In contrast, as histone-DNA interactions are strongest at the nucleosomal dyad, interfering with these contacts through acetylation of H3K122 reduces nucleosome stability (Tessarz and Kouzarides, 2014).

Another example of the ability of histone modifications to dramatically affect chromatin structure is citrullination of the linker histone H1 at arginine 54, a mark that has been recently identified as a modification specific to pluripotent cells. Citrullination describes the post-translational, enzymatic conversion of a positively charged arginine residue into citrulline, an uncharged, non-coded amino acid (Christophorou et al., 2014, Tessarz and Kouzarides, 2014). Due to charge neutralization, citrullination of arginine 54 interferes with the ability of histone H1 to bind DNA, thereby displacing the linker H1 from chromatin. As linker H1 is crucially important for the assembly of higher order chromatin structures, this histone modification contributes directly to the reduced chromatin condensation typical of the pluripotent cell state (Christophorou et al., 2014).

It is generally believed that chromatin remodellers apply their ability to reposition or eject nucleosomes to expose specific DNA sequences, including transcription factor binding sites and gene promoters (Clapier and Cairns, 2009). Nucleosomes that occlude gene promoters are major impediments for transcriptional initiation and must be displaced by ejection or repositioning for productive transcription (Lorch et al., 1987). The ability of chromatin remodellers to specifically remodel promoter nucleosomes in an activator-dependent manner can be reconstituted *in*

vitro, for example with budding yeast Chd1 at the *PHO5* gene promoter. Using a fully defined *in vitro* system, following multiple chromatographic steps, Chd1 was identified to be an essential factor for ejection of the promoter nucleosome at the repressed *PHO5* gene locus in response to the Pho4 activator. Notably, nucleosome remodelling was specific to the promoter immediately upstream of the transcription start site (TSS) and did not affect nucleosomes over the gene body. Crucially, this study also proved that chromatin remodellers are capable of acting in isolation to remodel the chromatin structure of promoter regions into a transcription competent state, without any requirements for additional histone modifying enzymes or histone chaperones (Ehrensberger and Kornberg, 2011).

Although histone modifications and chromatin remodelling are capable of independently manipulating chromatin structure, it is likely that they also act cooperatively. Indeed, many chromatin remodellers possess domains that recognize specific histone modifications, which may serve to recruit them to specific genes (Clapier and Cairns, 2009). For example, the pair of chromodomains in human CHD1 specifically recognizes the active H3K4me3 histone mark (Flanagan et al., 2005, Sims et al., 2005). Indeed, it has been observed that the recruitment of CHD1 to chromatin is enhanced by the H3K4me3 mark, thereby suggesting a plausible explanation for the localization of CHD1 at actively transcribed regions of the genome, including the de-condensed puff and interband regions of *Drosophila* polytene chromosomes (Stokes et al., 1996, Simic et al., 2003, Lin et al., 2011).

1.1.3 Determinants of Nucleosome Positioning

At a genome wide level, nucleosome positioning assumes a generalizable pattern – the TSS is surrounded by a nucleosome-free region (NFR) and flanked by highly positioned nucleosomes immediately upstream (-1) and downstream (+1) of the TSS. Nucleosomes are positioned with a regular periodicity in arrays extending away from both the -1 the +1 nucleosomes, though each successive nucleosome is less well positioned than the preceding one. Similarly, the transcription termination site (TTS) is also depleted of nucleosomes, though in contrast, the TTS is not

flanked by highly positioned nucleosomes. Incredibly, this stereotyped pattern of nucleosome positioning is conserved across eukaryotes, though the means by which it is achieved have, in some cases, diverged (Segal et al., 2006, Valouev et al., 2011, Struhl and Segal, 2013, West et al., 2014).

Reflecting the fact that histone-DNA nucleosomal contacts involve the DNA phosphate backbone, nucleosomes do not demonstrate traditional sequence specificity, such as that displayed by transcription factors (Luger et al., 1997, Struhl and Segal, 2013). Yet, the underlying DNA sequence can be a major determinant of nucleosome occupancy, arising from the requirement for nucleosomal DNA to bend around the histone octamer. Certain di-nucleotides, such as AA, AT or TA sequences, form a curvature in DNA; thus, the presence of such di-nucleotides at 10bp intervals are optimal for nucleosome formation since achieve curvature of the DNA sequence as a whole (Segal et al., 2006, Struhl and Segal, 2013).

Reconstitution of the nucleosomes *in vitro* with recombinant histones and yeast genomic DNA results in a pattern of nucleosome occupancy that correlates with the *in vivo* pattern. In fact, the DNA sequence alone ensures that up to 60% of genomic regions possess a nucleosome density comparable to that observed in the *in vivo* setting. Most notably, it is evident that the yeast DNA encodes for NFRs at the TSS and the TTS since these characteristic features are faithfully recapitulated *in vitro* (Kaplan et al., 2009, Zhang et al., 2009, Valouev et al., 2011, Zhang et al., 2011). This is probably due to the enrichment of homopolymeric poly(dA:dT) tracts at the promoters of yeast genes; due to their intrinsic stiffness, homopolymeric sequences prevent nucleosome occupancy (Iyer and Struhl, 1995, Struhl and Segal, 2013). Yet, by itself the DNA sequence is largely unable to specify the precise rotational positioning of nucleosomes; thus, a regular nucleosome periodicity extending into the gene body is not observed for nucleosomes reconstituted with yeast DNA and recombinant histone octamers (Kaplan et al., 2009, Zhang et al., 2009, Valouev et al., 2011, Zhang et al., 2011). However, incubation of the reconstituted chromatin with yeast extract in the presence of ATP is able to establish proper nucleosome positioning at the 5' end of genes. As this process is strictly ATP dependent, it is probably mediated by ATP-dependent chromatin remodellers, though a role for enzymes such as kinases cannot be formally excluded (Zhang et al., 2011). Indeed,

genome-wide analyses of nucleosome occupancy and DNA accessibility in *Drosophila* following depletion of various chromatin remodellers suggest that chromatin remodelling enzymes actively reposition nucleosomes, overriding their intrinsic sequence preferences. In this manner, NuRD, (P)BAP and INO80 complexes increase nucleosome density on sequences that are intrinsically unfavourable to nucleosomes; in contrast, ISWI reduces nucleosome occupancy at its target sites, which are generally intrinsically favourable to nucleosome formation (Moshkin et al., 2012).

An integral role for chromatin remodellers in nucleosome positioning at a genome-wide level is further supported by analyses of nucleosome positions in yeast strains lacking specific chromatin remodellers. For example, the chromatin remodellers ISW1b, INO80, CHD1 and SWI/SNF shift nucleosomes away from NFRs (i.e. nucleosomes downstream of the TSS are moved in the 5' direction, whereas nucleosomes immediately upstream of the TTS are moved in the 3' direction). In contrast, ISW1a and ISW2 move nucleosomes towards NFRs (Yen et al., 2012). Due to biological redundancy, the absence of individual chromatin remodellers does not result in significant alterations in nucleosome positioning; however, a budding yeast strain (*isw1Δ, chd1Δ*) lacking the CHD1 and ISW1 remodellers exhibits compromised nucleosome positioning in the gene body, though the +1 promoter nucleosome remains unaffected (Gkikopoulos et al., 2011). This function appears to be highly conserved since the evolutionarily distant *Schizosacchromyces pombe*, which does not encode for any members of the ISWI family, relies upon its CHD1 homologues, Hrp1 and Hrp3, to maintain proper nucleosome positioning in the gene body (Flaus et al., 2006, Pointner et al., 2012, Shim et al., 2012). Curiously, however, depletion of Chd1 from mouse embryonic fibroblasts (MEFs) results in a global reduction in nucleosome occupancy, though most prominently upstream and downstream of the TSS, without affecting nucleosome positioning (Skene et al., 2014). Thus, while the overall principles of nucleosome positioning appear to be conserved across eukaryotes, it is probable that they differ at a precise, mechanistic level (Struhl and Segal, 2013).

Indeed, much of our understanding of the determinants of nucleosome positioning is based on studies of *Sacchromyces cerevisiae*. Although many facets can

probably be extrapolated to higher eukaryotes, important inter-species differences must be recognized. Most importantly, HHO1p, the budding yeast homologue of histone H1, is present at very low levels, whereas the histone H1 is widely incorporated into the chromatin of higher eukaryotes, promoting the higher order packaging (Struhl and Segal, 2013). Additionally, the nucleosomal architecture around gene promoters differ slightly between budding yeast and metazoans – while the TSS is located within the +1 nucleosome in the former, it is typically situated in the nucleosome depleted region in metazoans. Likewise, the average spacing between nucleosomes displays inter-species variation (Segal et al., 2006, Mavrich et al., 2008, Schones et al., 2008, Lantermann et al., 2010, Li et al., 2011, Valouev et al., 2011, Skene et al., 2014). These idiosyncrasies probably reflect the specific complement of chromatin remodellers present within each organism in addition to differences in transcriptional activity (Struhl and Segal, 2013).

1.1.4 Histone Modifications

Besides targeting histone residues that contribute to histone-DNA contacts (as has been discussed previously – refer to subsection **1.1.2**), post-translational modifications (PTMs) also frequently involve the tails of the core histones, which are accessible as they are disordered regions that extend away from the globular histone core. In general, covalent histone modifications can be regarded to perform one of two functions: 1) to directly affect the chromatin structure, for example by destabilizing nucleosomes, or 2) to act as signals for the recruitment of other protein factors.

The ability of acetylation of specific residues of histone H3 (e.g. lysines 56 and 122) to affect chromatin structure has been mentioned. In brief, however, as the lysine residues targeted by acetylation are often important histone-DNA contacts, neutralization of the positive charges of lysine residues by acetylation contributes to the destabilization of nucleosomes and in some cases, prevents the formation of higher order chromatin structures (Watanabe et al., 2010, Tessarz and Kouzarides, 2014). Although acetylation of lysine 16 of histone H4 (H4K16ac) does not appear to affect histone-DNA contacts, this histone modification reduces the propensity of

nucleosomal array to self-associate, thus compromising chromatin compaction (Shahbazian and Grunstein, 2007). Additionally, H4K16ac acts as a boundary element to prevent the spreading of heterochromatin (Suka et al., 2002). Thus, heterochromatin spreading requires that the Sir2 histone deacetylase – a component of the silent information regulator (Sir) complex that is a major constituent of heterochromatin – de-acetylates histone 4 at lysine 16, such as to promote the binding of additional molecules of the Sir complex, which has a higher affinity for hypo-acetylated histone H4 tails, to adjacent regions of chromatin (Shahbazian and Grunstein, 2007). (The other main occurrence of histone acetylation is on newly synthesized histones, though the acetyl moiety is rapidly removed by histone deacetylases shortly after deposition of these histones into chromatin. In many eukaryotes, the main sites of acetylation on histone H4 are lysines 5 and 12; acetylation sites on histone H3 are more variable – lysines 9 and 14 in *Tetrahymena* and lysines 14 and 23 in *Drosophila* (Sobel et al., 1995, Shahbazian and Grunstein, 2007).)

In contrast, PTMs that do not affect the net charge, such as lysine methylation for instance, are unlikely to directly affect nucleosome stability by interfering with histone-DNA contacts. Instead, these histone modifications are likely to act by coordinating the recruitment of a diverse set of factors to chromatin. This is feasible because histone mark-reading domains are capable of distinguishing between histone marks with a high degree of precision. In fact, it appears that each PTM can be recognized several classes of histone mark-reading domains. For example, an incomplete list of motifs that recognize methylated lysine residues includes plant homeodomain (PHD) fingers, bromo-adjacent homology domains and most prominently, “Royal family” motifs – a family that encompasses Tudor, PWWP domains and chromodomains (Patel and Wang, 2013).

An example of canonical methyl-lysine recognition is the interaction between the chromodomain of heterochromatin protein 1 (HP1) and histone H3 tails modified by tri-methylation of lysine 9. The tail of histone H3 makes extensive β-sheet interactions with two β-strands of the HP1 chromodomain to form an antiparallel β-barrel (Figure 1.4B). Crucially, the methyl-lysine side chain, occupies a surface groove pocket that culminates in a three-sided aromatic cage (Figure 1.4A). These

three conserved aromatic residues stabilize the methyl-lysine by cation-π interactions (Jacobs and Khorasanizadeh, 2002). Comparable aromatic cage architectures have adopted by other "Royal family" motifs to recognize methylation at other residues on the histone H3 tail, for example, by the chromodomain of *Drosophila* Polycomb to bind H3K27me3, and by the Tudor domain of PHF1 to recognize H3K36me3 (Min et al., 2003, Musselman et al., 2012a).

Figure 1.4 – Recognition of the H3K9me3 Modification by the HP1 Chromodomain
A. The crystal structure of the chromodomain of HP1 in complex with an H3K9me3 peptide reveals that the histone methyl-lysine side chain occupies a surface groove that has a three-sided aromatic cage at one end. These three conserved aromatic residues stabilize the methyl-lysine by cation-π interactions.
B. The histone H3 tail also engage in extensive β-sheet interactions with two β-strands of the HP1 chromodomain to form an antiparallel β-barrel.
The images were taken from (Patel and Wang, 2013).

Recognition of the histone acetyl-lysine modification is mediated by bromodomains, which organize into anti-parallel four-helix bundles linked by variable loop regions (Figure 1.5A). The four helices establish a deep binding pocket that is primarily hydrophobic in character. Extending beyond the base of the helical bundle, the loop regions deepen the binding pocket, forming its rim. Due to the variable nature of the loop regions, they allow individual bromodomains to acquire unique surface characteristics that dictate binding specificities (Filippakopoulos et al., 2012, Patel and Wang, 2013). The acetyl-lysine side chain projects deep into the binding pocket, with the hydrophobic walls of the pocket packing against, and thus stabilizing, the aliphatic side chain. The specificity of bromodomains for acetyl-lysine is heavily influenced by the electrostatic potential of binding pocket – without the charge-neutralizing acetyl PTM, an interaction between the positively charged lysine residue and the hydrophobic pocket is energetically unfavourable. A conserved asparagine residue is also a key determinant of bromodomain acetyl-lysine recognition since the amide of the conserved asparagine forms a crucial hydrogen bond with the oxygen of the carbonyl group of the acetyl moiety (Owen et al., 2000, Patel and Wang, 2013). This intermolecular hydrogen bond is critically important as it serves to position the acetyl-lysine in the correct orientation. The bromodomain of the transcriptional co-repressor, KAP1, lacks this conserved

asparagine; correspondingly, it is unable to recognize the acetyl-lysine modification (Zeng et al., 2008).

Figure 1.5 – Recognition of the H4K16ac Modification by the Gcn5p Bromodomain
A. Bromodomains mediate histone acetyl-lysine recognition, as demonstrated in this structure of the Gcn5p bromodomain in complex with an H4K16ac peptide. A deep, largely hydrophobic binding pocket is created at one end of the four-helix bundle of the bromodomain. The acetyl-lysine side chain projects deep into the binding pocket. A conserved bromodomain asparagine residue forms a crucial hydrogen bond with the oxygen of the carbonyl group of the acetyl moiety.
B. The hydrophobic walls of the binding pocket pack against, and thus stabilize, the aliphatic acetyl-lysine side chain. Additionally, the acetyl moiety is anchored to a bromodomain asparagine residue by hydrogen bonding.
The images were taken from (Patel and Wang, 2013).

Actively transcribed genes can typically be distinguished from repressed genes on the basis of their histone modifications. In addition to hyper-acetylation of histones H3 and H4, the promoters of active genes are typically associated with the H3K4me3 modification, which is deposited by members of the Trithorax family of lysine methyltransferases (Santos-Rosa et al., 2002, Kouzarides, 2007, Suganuma and Workman, 2011). Conversely, regions of constitutive heterochromatin that always remain highly compacted and transcriptionally inactive tend to be marked by H3K9me3. In contrast, facultative heterochromatin – repressed regions that can, nonetheless, be induced to decondense and permit transcription in a context-specific manner – are enriched in H3K27me3 (Trojer and Reinberg, 2007, Suganuma and Workman, 2011). Hence, H3K4me3 could, simplistically, be regarded as an ‘active’ mark, in contradistinction to H3K9me3 and H3K27me3, which are prototypical ‘repressive’ marks.

Although histone modifications probably do not represent a ‘code’ akin to the genetic code, there is emerging evidence that they can directly influence transcriptional activity. Most notably, H3K4me3 binds directly to the general transcription factor, TFIID, via the PHD domain of the TAF3 subunit (Vermeulen et al., 2007). Moreover, employing a cell-free transcription assay, it was demonstrated that the H3K4me3-TAF3 interaction directly recruits TFIID to promoters and facilitates the assembly of the transcription pre-initiation complex. Importantly, recruitment of TFIID by H3K4me3 is able to compensate for the absence of a

functional TATA box as assessed by activator-dependent transcriptional activity *in vitro* (Lauberth et al., 2013). Although the promoter typically resides in the NFR, it is possible that H3K4me3-TAF3 interaction promotes pre-initiation complex formation by recruiting TFIID to the peak of H3K4me3 just downstream of the TSS, thereby ensuring an accumulation of TFIID in the general vicinity of the promoter.

H3K9me3 is essential for the maintenance of heterochromatin. Deletion of Clr4, the H3K9me3 lysine methyltransferase in fission yeast, or its two Suv39h homologues in mammals, results in defects in pericentric and telomeric heterochromatin, in addition to compromised gene silencing (Nakayama et al., 2001, Peters et al., 2003, Garcia-Cao et al., 2004). This histone mark achieves transcription repression through the recruitment of HP1 to chromatin, via a chromodomain-mediated interaction with H3K9me3 (Kouzarides, 2007). Indeed, enforced recruitment of HP1 to euchromatic gene loci is sufficient to achieve transcriptional repression though a process that ultimately results in local chromatin compaction (Ayyanathan et al., 2003). Moreover, it has been proposed that the H3K9me3 modification can be propagated – an initial H3K9me3 mark that serves as the nucleation site interacts with HP1, which in turn, causes the recruitment of a Suv39h methyltransferase to chromatin, such that the lysine 9 residue of neighbouring histone H3 tails is methylated, resulting in yet more HP1 binding sites (Hall et al., 2002, Grewal and Moazed, 2003, Muramatsu et al., 2013). Indeed, the dosage-dependent role of HP1 in heterochromatic silencing might imply that such a self-amplifying positive feedback loop could be essential for the formation and maintenance of constitutive heterochromatin (Grewal and Moazed, 2003).

The genes targeted by the H3K27me3 mark are frequently associated with differentiation and key developmental processes. Hence, in a developmental stage-dependent and tissue-specific manner, regions of facultative chromatin decondense to permit transcription (Trojer and Reinberg, 2007, Schwartz and Pirrotta, 2013). The transcription repression of these Polycomb target genes involves the two PRC complexes. Containing the histone methyltransferase, Ezh2, the PRC2 complex tri-methylates histone H3 at lysine 27. It is suggested that PRC1 is then recruited to chromatin through the chromodomains of its Cbx subunits, which display an affinity for H3K27me3. In this indirect manner, RING1 and RING2, the E3 ubiquitin ligase subunits of PRC1, are thereby brought into close proximity

to their target for mono-ubiquitylation, histone H2A (Schwartz and Pirrotta, 2013). Mono-ubiquitylated H2A (H2Aub) represses transcription by interfering with the recruitment of transcription elongation factors, such as the FACT complex (Zhou et al., 2008). Additionally, H2Aub mediates transcriptional repression at the level of higher order chromatin structure, as reconstituted nucleosomes bearing H2Aub interact more readily with the linker histone H1 than unmodified nucleosomes (Jason et al., 2005, Weake and Workman, 2008). Conversely, it is also suggested that de-ubiquitylation of H2A promotes dissociation of H1 from nucleosomes (Zhu et al., 2007, Weake and Workman, 2008). PRC1 may also mediate transcriptional repression in an H2A ubiquitylation-independent manner, perhaps by promoting chromatin compaction (Eskeland et al., 2010, Grau et al., 2011).

In contrast to H2Aub, ubiquitylated H2B (H2Bub) is regarded as an active mark. Ubiquitylation of H2B is performed by the E2 ubiquitin conjugating enzyme, Rad6, and the E3 ubiquitin ligase, Bre1 (Weake and Workman, 2008). Requirements for ubiquitylation include active transcription and the presence of the PAF complex, a factor associated with RNAPII. Indeed, Rad6 directly interacts with elongating RNAPII (Wood et al., 2003, Xiao et al., 2005). Significantly, H2Bub is required for di- and tri-methylation of histone H3 at lysines 4 and 79, which are important active marks (as discussed earlier). Therefore, transcriptional repression is frequently associated de-ubiquitylation of histone H2B, which in budding yeast, is performed by Ubp8, a component of the SAGA complex, and Ubp10. Curiously, Ubp8 and Ubp10 are synergistic, and probably act on different pools of H2Bub (Weake and Workman, 2008). In *Drosophila*, biochemical and genetic evidence implicate the GMPS/USP7 complex as a transcriptional repressor as a result of its selective H2B de-ubiquitylation activity. The GMPS/USP7 complex participates not only in Polycomb silencing of homeotic genes – mutations in either protein result in homeotic transformations – but also in the repression of ecdysone target genes – hormonally-regulated genes that are essential for normal development (van der Knaap et al., 2005, van der Knaap et al., 2010). Histone ubiquitylation is also employed as a signaling mechanism for DNA repair. The E3 ubiquitin ligases RNF8 and RNF168 ubiquitylate histone H2A around sites of damage enabling the ubiquitylation-dependent recruitment of DNA repair effectors, such as 53BP1 and RAP80, to chromatin (Panier and Durocher, 2013).

1.1.5 Differentiation as a Paradigm of Chromatin State Transitions

The chromatin architecture not only reflects, but also abets the transcriptional activity of a cell. Thus, transitions in cellular state, which involve the adoption of distinct transcriptional programmes, correspondingly, are accompanied by significant and distinctive alterations in chromatin state. Characterized by the ability to differentiate into any cell arising from the three germ layers, pluripotency is indisputably one of the most unique of cell states. Regulated by a set of master regulators including the transcription factors, Oct4, Nanog and Sox2, the pluripotent cell state has been primarily studied in embryonic stem cells (ESCs), which are cells derived from the inner cell mass of blastocysts, exhibit indefinite self-renewal, and are amenable to *ex vivo* tissue culture (Boyer et al., 2005, Wang et al., 2006, Takahashi et al., 2007, Chen et al., 2008). Unsurprisingly, it has become apparent that the chromatin structure of ESCs is highly distinctive.

ESCs are enriched in loosely packaged euchromatin and lack the abundant, punctate distribution of highly compact heterochromatic foci typical of differentiated cells (Meshorer et al., 2006). Supporting the notion of a globally open, transcriptionally permissive chromatin environment in ESCs are the observations that histones – both the core histones and the linker histone, H1 – associate less avidly with ESC chromatin and that histone exchange occurs at a faster rate in ESCs. Indeed, this permissive chromatin environment is reflected in a global increase in transcriptional activity; for example, higher levels of transcription at intronic and intergenic regions of the genome have been observed in ESCs (Efroni et al., 2008).

Although approximately 80% of promoters in ESCs are marked by the active H3K4me3 histone modification, in the region of 22% of these concomitantly display the repressive H3K27me3 mark – a feature termed bivalency (Bernstein et al., 2006, Mikkelsen et al., 2007). The genes that feature these bivalent marks are not actively transcribed in ESCs and are disproportionately over-represented by regulators of differentiation and developmental processes. Thus, it was postulated that bivalent marks keep genes encoding for important developmental regulators in a poised but nonetheless, transcriptionally repressed state (Azuara et al., 2006,

Bernstein et al., 2006). The vast majority of bivalent marks are resolved during differentiation by the loss of one of the two opposing histone modifications, and is accompanied by either transcriptional activation or persistent gene repression (Mikkelsen et al., 2007, Gifford et al., 2013, Xie et al., 2013). It is currently believed that bivalent marks serve to increase the precision with which gene expression is controlled during development – having more input sources enhances the robustness of the regulatory network controlling the transcription of important developmental genes and minimizes the possibility of inopportune gene expression (Voigt et al., 2013).

In addition to the resolution of bivalent marks, differentiation of ESCs is associated with widespread deposition of the H3K27me3 mark across the genome (Figure 1.6) (Hawkins et al., 2010, Zhu et al., 2013). (A similar genome-wide increase in the H3K9me3 mark has subsequently been attributed to being an artifact of tissue culture rather than an effect of differentiation *per se* (Zhu et al., 2013).) Thus, the permissive chromatin state in ESCs is replaced during differentiation by a restrictive and compacted chromatin environment dominated by widespread Polycomb-mediated repression (Hawkins et al., 2010, Zhu et al., 2013).

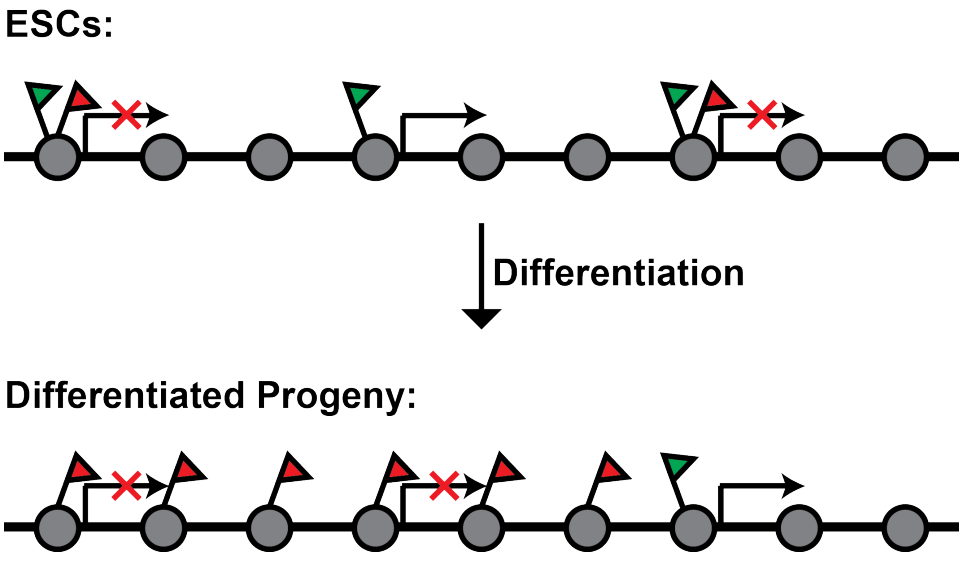


Figure 1.6 – Differentiation of ESCs is accompanied by a Chromatin State Transition
This schematic depicts the changes in chromatin structure associated with the differentiation of pluripotent ESCs. Green flags represent the active H3K4me3 mark, while red flags denote the repressive H3K27me3 modification. ESC chromatin is uniquely open and many silent genes encoding for developmental regulators possess bivalent marks. Differentiation is associated with the resolution of bivalent marks, progressive chromatin

compaction, and widespread accumulation of the repressive, Polycomb-mediated H3K27me3 modification across the genome.

In contrast to the large-scale changes in histone modifications, however, differentiation of ESCs does not result in widespread changes in nucleosome occupancy (Teif et al., 2012, West et al., 2014). Instead, most differences in nucleosomal occupancy between ESCs and their differentiated progeny involve small genomic regions the approximate size of mononucleosomes. Nonetheless, these regions of variable nucleosome occupancy are enriched for the DNA binding sites of factors involved in the regulation of pluripotency. Overall, however, in accordance with the notion of an increase in chromatin compaction upon differentiation, the majority of sites of variable nucleosome occupancy are associated with lower occupancy in ESCs than in their differentiated counterparts (West et al., 2014). Therefore, while various chromatin remodellers have been implicated as important determinants in the maintenance of pluripotency, it is unlikely that they function by mediating dramatic alterations in nucleosomal occupancy (Gaspar-Maia et al., 2009, Ho et al., 2009b). Instead, it is more probable that they act by regulating the transcription of key pluripotency or developmental genes.

1.2 Chromatin Remodelling Enzymes

As chromatin remodellers share many similar characteristics, a panoramic survey of these enzymes is of considerable utility. Nevertheless, where possible, general principles will be illustrated with examples pertaining to CHD proteins and the SMARCAD1-KAP1 complex, since they are the main focus of this study.

1.2.1 Classification of Chromatin Remodellers

Chromatin remodellers are enzymes that couple the energy released from ATP hydrolysis to repositioning, ejecting, or restructuring nucleosomes. Although a large and diverse group of proteins, members of the SWI2/SNF2-like family are unified in their shared possession of a helicase-like domain formed by a core of two recA-like domains (Durr et al., 2005, Clapier and Cairns, 2009). Multiple sequence alignment

of SWI2/SNF2-like family members identifies twenty-four distinct subfamilies that can be classified into six groups: 1) Snf2-like, 2) Swr1-like, 3) SSO1653-like, 4) Rad54-like, 5) Rad5/16-like and 6) distantly-related remodellers (Flaus et al., 2006).

The Snf2-like group includes several notable subfamilies, including the SWI/SNF, ISWI, CHD1 and Mi2 (CHD3/CHD4) subfamilies (Flaus et al., 2006). Eukaryotes typically have at least two chromatin remodelling complexes that have at their cores, members of the SWI/SNF subfamily. For instance, *Sacchromyces cerevisiae* possesses the SWI/SNF and RSC complexes, whose ATPase activities are conferred by the paralogues Swi2/Snf2 and Sth1 respectively (Flaus et al., 2006, Clapier and Cairns, 2009). Typically, amongst the subunits of these complexes are nuclear actin-related proteins, which interact via a helicase-SANT-associated (HSA) domain on the core ATPase subunit (Szerlong et al., 2008, Clapier and Cairns, 2009). The presence of multiple bromodomains, which facilitate recognition of acetyl-lysine residues on histone tails, is another characteristic feature of SWI/SNF chromatin remodelling complexes – for instance, eight of the fifteen bromodomains encoded by budding yeast are distributed amongst the fifteen subunits of the RSC complex, including the catalytic Sth1 subunit (Cairns et al., 1996, Clapier and Cairns, 2009). Biochemically, the essential RSC complex has 3'→5' translocase activity, slides and ejects nucleosomes (Lorch et al., 1999, Saha et al., 2002, Lorch et al., 2006).

C-terminal SANT and SLIDE domains, which contribute to DNA binding, are distinguishing features of ATPases of the ISWI subfamily (Grune et al., 2003, Clapier and Cairns, 2009, Yamada et al., 2011). (The SANT domain is related to Myb DNA-binding domain and its name reflects its frequent presence in proteins associated with chromatin, including Swi3, Ada2, N-CoR and TFIIB (SANT). The similar SLIDE domain stands for SANT-like ISWI domain (Grune et al., 2003, Boyer et al., 2004).) Many complexes of ISWI subfamily members, including budding yeast Isw1a and *Drosophila* ACF (ATP-utilizing chromatin assembly and remodelling factor), centre mononucleosomes reconstituted on short DNA fragments and mediate regulate spacing of nucleosomal arrays *in vitro* (Ito et al., 1999, Tsukiyama et al., 1999, Stockdale et al., 2006). Indeed, supporting a general role in regulating nucleosome spacing, a *S. cerevisiae* strain (*isw1Δ, chd1Δ*)

lacking both the CHD1 and ISW1 remodellers experiences a genome-wide loss of proper nucleosome positioning in the gene body (Gkikopoulos et al., 2011). Likewise, in *Drosophila*, ISWI acts across the genome to reduce nucleosome density at sequences that are intrinsically favourable to nucleosome occupancy (Moshkin et al., 2012).

As suggested by their name, chromodomain, helicase, DNA-binding (CHD) chromatin remodellers are characterized by an N-terminal tandem chromodomain. Although conserved from yeast to man, a single CHD protein in present in budding yeast, Chd1, as compared to the nine (CHD1-9) in mammals (Marfella and Imbalzano, 2007, Clapier and Cairns, 2009). CHD proteins can, however, be further sub-divided into three sub-families, Chd1 (the prototype of which is budding yeast Chd1, but which also includes mammalian CHD1 and CHD2), Mi2 (featuring CHD3-5), and CHD7 (a subfamily encompassing the relatively uncharacterized CHD7-9 proteins) (Flaus et al., 2006). The DNA-binding domain is actually specific to the Chd1 subfamily and is structurally homologous to the SANT and SLIDE domains of ISWI remodellers, though they share limited sequence homology (Ryan et al., 2011, Sharma et al., 2011). Indeed, it has become increasingly appreciated that Chd1 and remodellers of the ISWI subfamily share numerous similarities, not only in terms of their structural features, but also in their catalytic properties – Chd1 also centres mononucleosomes and promotes the regular spacing of nucleosomal arrays (Lusser et al., 2005, Stockdale et al., 2006, Patel et al., 2011). Strikingly, Isw1 and Chd1 display functional redundancy in budding yeast – while strains bearing individual deletions display only mild phenotypes, double mutants exhibit synthetic phenotypes, including a defect in global nucleosome positioning (Gkikopoulos et al., 2011).

Lacking SANT and SLIDE domains, remodellers of the Mi2 subfamily have tandem PHD domains in their N-termini instead. The tandem PHD domains mediate recognition of histone tails; for example, those of CHD4 bind the H3K9me3 modification, whereas the CHD5 PHD domains are specific for unmodified H3 tails (Mansfield et al., 2011, Paul et al., 2013). CHD3 and CHD4 are the catalytic subunits of the multi-subunit nucleosome remodelling and histone deacetylation (NuRD) complex (Tong et al., 1998, Xue et al., 1998, Zhang et al., 1998, Wade et

al., 1999). As might be suggested from its ability to deacetylate histones, the role of the NuRD complex is that of a transcriptional repressor (Clapier and Cairns, 2009).

The Swr1-like group is characterized by a “split” ATPase domain – an architecture that arises from the long insert separating the two recA-like lobes of the ATPase domain. The Swr1-like group of remodellers includes two sub-families of general importance, Ino80 and Swr1 (Flaus et al., 2006, Clapier and Cairns, 2009). The long insert serves as a platform to recruit regulatory subunits, such as the AAA+ ATPases, Rvb1 and Rvb2, to the catalytic subunit (Wu et al., 2005, Nguyen et al., 2013, Tosi et al., 2013). In addition to nucleosome sliding, the SWR1 and the INO80 complexes catalyse histone exchange (Morrison and Shen, 2009). SWR1 replaces conventional H2A/H2B dimers with free variant H2AZ/H2B dimers, while INO80 performs the converse reaction, exchanging incorporated H2AZ/H2B dimers for free H2A/H2B dimers (Mizuguchi et al., 2004, Papamichos-Chronakis et al., 2011).

1.2.2 General Mechanism of Chromatin Remodellers

Related as they are to helicases, it is consequently unsurprising that SWI2/SNF2-like chromatin remodelling ATPases are capable of processive translocation on DNA (Saha et al., 2002, Clapier and Cairns, 2012). It has been observed that ATPase domains of both ISWI and SWI/SNF remodellers bind nucleosomal DNA approximately 20bp, or two superhelical turns, away from the dyad axis (i.e. superhelix location (SHL) -2/+2) (Saha et al., 2005, Dang and Bartholomew, 2007). Thus, one popular model suggests that a chromatin remodeller would anchor itself to the nucleosome through its ATPase domain, which binds nucleosomal DNA at SHL +2 or -2. While the ATPase domain remains affixed to the nucleosome at that invariant position, its DNA-binding domain draws linker DNA into the nucleosome, creating a DNA loop on the nucleosome surface. The ATPase domain is believed to then pump the DNA along the nucleosome, towards the dyad. Presumably, each power stroke only breaks one or two DNA-histone contacts; however, by repeating this process, the transient DNA loop may be resolved by propagation around the nucleosome (Clapier and Cairns, 2009). While intuitive, it is unclear whether this

model of nucleosome sliding is compatible with chromatin remodellers that, for example, lack DNA-binding domains.

Figure 1.7 – Model of Nucleosome Sliding by Addition of DNA
Linker DNA is drawn into the nucleosome by a DNA-binding domain (annotated as DBD in the figure), creating a DNA loop on the nucleosome surface (State 2). The ATPase domain then pumps DNA towards the nucleosomal dyad by breaking DNA-histone contacts and allowing the DNA to propagate around the nucleosome, thus resolving the DNA loop (State 3). If the DNA-binding domain rebinds to linker DNA, this cycle can then be repeated. This image was taken from (Clapier and Cairns, 2009).

However, the aforementioned model is disfavoured by data from a recent study of ISWI proteins employing a single-molecule, fluorescence resonance energy transfer technique (FRET). It was observed that the first phase of the nucleosome sliding process was the exit of 7bp of nucleosomal DNA, accomplished by extruding DNA 1bp at a time. 3bp of DNA was then drawn into the nucleosome, presumably because the initial removal of 7bp of DNA generates tension at the DNA entry side. This cycle could then be repeated by extruding another 3bp of DNA in single base pair increments, such as to accumulate the 7bp deficit in nucleosomal DNA content required for triggering DNA entry (Figure 1.8) (Deindl et al., 2013).

Figure 1.8 – Model of Nucleosome Sliding by Removal of DNA by ISWI
The first step of the nucleosome sliding process by ISWI is the exit of 7bp of nucleosomal DNA, accomplished by the extrusion of DNA in 1bp steps. This generates tension at the DNA entry site, resulting in 3bp of DNA being drawn into the nucleosome. This cycle may then be repeated by extruding another 3bp of DNA in single base pair increments to accumulate the 7bp deficit in nucleosomal DNA content required to trigger DNA entry. The image was taken from (Deindl et al., 2013).

This model, however, is perplexing since it implies that during the repositioning reaction, nucleosomes have to accommodate a deficit in DNA content of between 4 to 7bp (Narlikar et al., 2013). Indeed, the crystal structure of the nucleosome does not suggest laxity in the way nucleosomal DNA is wound around the histone octamer (Luger et al., 1997). Although DNA could be underwound, for instance at SHL2 and SHL5, to cope with a 1bp reduction in nucleosomal DNA content, it remains difficult to conceive of the structural basis through which nucleosomes can accommodate a deficit of up to 7bp (Narlikar et al., 2013).

As these models are contradictory – one implies that DNA is first drawn into the nucleosome, while the other suggests that DNA is first extruded out of the nucleosome – it is likely that they are employed in a mutually exclusive manner by different chromatin remodellers. Thus, additional characterization of a diverse range of chromatin remodellers is required to determine which of the nucleosome sliding mechanisms is adopted by each specific remodeller. Much of the existing mechanistic characterization of chromatin remodellers has focused upon complexes with relatively few subunits, such as ISWI; thus, it would be informative to assess whether larger complexes, such as RSC, SWR1 or INO80, share a conserved mechanism for sliding nucleosomes. It is also remains unclear how chromatin remodellers couple ATP hydrolysis to the breaking of DNA-histone contacts. For Chd1, catalysis is associated with a conformational change in its ATPase domain; however, the ATPase motor is not linked to the DNA-binding SANT and SLIDE domains by a rigid connector through which force can be transmitted to push DNA out of the nucleosome (Nodelman and Bowman, 2013). Thus, further mechanistic studies are required to establish our understanding of nucleosome sliding by chromatin remodellers, and to elucidate the mechanisms behind other activities such as nucleosome ejection or histone dimer exchange.

1.2.3 Regulation of Chromatin Remodellers

Complementing the conserved ATPase domain, chromatin remodellers typically possess additional accessory domains. Through the recognition of substrates such as DNA and histone tails, these accessory domains act as both positive and negative regulators. Arguably the most highly characterized accessory domain is the HAND-SANT-SLIDE (HSS) domain that was first identified in the *Drosophila* ISWI protein and mediates binding of linker DNA (Grune et al., 2003, Dang and Bartholomew, 2007). Although the HSS domain significantly increases the affinity of the chromatin remodeller for nucleosomes, a fragment of the ISWI protein spanning only the core ATPase domain is nonetheless capable of sliding nucleosomes, albeit at a reaction rate an order of magnitude slower than that of the full-length protein (Grune et al., 2003, Mueller-Planitz et al., 2013). Within the context of the *S. cerevisiae* ISW2 complex in its entirety, however, deletion of the

SLIDE domain does not drastically reduce the remodeller's affinity for either DNA or nucleosomes. Yet, disproportionately greater defects in nucleosome-stimulated ATPase and nucleosome sliding activities were associated with the ISW2 complex incorporating the Isw2 mutant lacking its SLIDE domain, suggesting that the LSIDE domains may contribute to processes beyond DNA binding (Hota et al., 2013).

A role for the HSS domain as a positive regulator of nucleosome recognition and sliding is not solely confine to ISWI-type chromatin remodellers. A structurally homologous SANT and SLIDE domain has been identified in Chd1. Correspondingly, deletion of these domains in Chd1 compromises its affinity for nucleosomes and its ability to slide nucleosomes (Ryan et al., 2011, Sharma et al., 2011). Moreover, the Chd1 SANT and SLIDE domains also appear to dictate the directionality of nucleosome sliding (McKnight et al., 2011). Other domains that modulate the nucleosome recognition capability of chromatin remodellers may have a comparable effect on catalysis. Indeed, the tandem PHD domains enhance the nucleosome sliding activity of CHD4, albeit when tested under the slightly non-physiological setting of free CHD4 lacking additional subunits (Watson et al., 2012).

Modular allostery – inhibition of the catalytic domain by another distinct structural domain of the same enzyme – is a strategy of negative regulation adopted by some chromatin remodellers. This mode of regulation is exemplified by the organization of the ATPase domain of Chd1 relative to its tandem chromodomains. The tandem Chd1 chromodomains are connected by two helices that pack against each other, forming a wedge-like structure. This 'chromo-wedge' occupies the central cleft separating the two RecA lobes of the ATPase domain, presumably stabilizing it in an open, inactive conformation (Figure 1.9). This form of conformational modular allostery is effective since the 'chromo-wedge' physically impedes the second ATPase lobe from rotating the fifty-two degrees required to adopt a closed, active configuration that permits ATP hydrolysis. Moreover, the 'chromo-wedge' has a highly acidic character that allows it to bind a basic surface on the second ATPase lobe that acts as a conserved DNA-binding interface. Thus, by inhibiting DNA and nucleosome recognition, the chromodomains also negatively regulate ATPase activity of Chd1 through steric modular allostery (Hauk et al., 2010).

Figure 1.9 – Modular Allostery by the CHD1 Chromodomains
A. The chromodomains of CHD1 occupy the cleft between the two ATPase lobes, stabilizing them in an open, inactive configuration. Additionally, the chromo-wedge, consisting of the two helices that separate the chromodomains, packs against the DNA-binding surface of the ATPase motor, thereby preventing nucleic acid binding.
B. This schematic illustrates how the ATPase domain of CHD1 is arranged in an open state. The second lobe must undergo a 52° rotation for ATPase motor to assume the active, closed conformation seen in the ATPase motor of the Vasa helicase.
The images were taken from (Hauk et al., 2010).

The *Drosophila* ISWI protein is also highly regulated by modular allostery through two auto-inhibitory domains, AutoN and NegC. AutoN antagonizes ATPase activity by competing with the histone H4 tail for binding to a region on ISWI; correspondingly, deletion of AutoN relieves ISWI of its dependence on the H4 tail for ATPase activity. In contrast, NegC interferes with the coupling of ATPase activity with DNA translocation. It is believed that binding of the HSS domain to extranucleosomal DNA relieves the auto-inhibition of NegC, thereby ensuring that productive nucleosome sliding is limited to nucleosomes flanked by linker DNA of a sufficient length (Clapier and Cairns, 2012). Complementary single-molecule FRET analysis of human ACF, a chromatin remodelling complex of the ISWI subfamily, has confirmed the auto-inhibitory function of the AutoN domain of the Snf2h catalytic subunit. However, in the context of the entire ACF complex, NegC did not influence the relationship between linker DNA length and nucleosome sliding activity. Significantly, in addition to conferring H4 tail dependence on ACF ATPase activity, AutoN also serves as a sensor for the length of linker DNA – deletion of AutoN increases the rate at which ACF remodels nucleosomes with limited linker DNA. It is suggested that for nucleosomes with short linker DNA regions, the accessory subunit, Acf1, binds predominantly to the histone H4 tail, thereby preventing the H4 tail from relieving AutoN-mediated auto-inhibition. In contrast, in the context of increased linker DNA length, Acf1 instead binds preferentially to linker DNA, leaving the histone H4 tail free to displace AutoN from the ATPase domain, stimulating ATPase activity (Hwang et al., 2014).

It is clear, therefore, that auto-inhibitory accessory domains are conserved features of Chd1-like and ISWI-like chromatin remodellers. Yet, as the relatively distant Rhp26 protein, the *S. pombe* homologue of CSB (an enzyme that is crucial for transcription-coupled nucleotide excision repair), also possesses a functionally

comparable repressive (leucine-latch) domain, it is likely that auto-inhibition by modular allostery is a common regulatory feature of chromatin remodellers (Wang et al., 2014a).

Maintaining chromatin remodellers in an inactive state by default may be a general regulatory mechanism for cells for several reasons. First, as the human genome encodes for thirty-three proteins that possess a SWI2/SNF2-like helicase domain, it is likely that they are specialized for specific biological processes, with the majority not normally being required. Thus, maintaining a reserve of inactive chromatin remodellers that can be rapidly converted into an active conformation is a simple yet effective means by which cells can dynamically modify their chromatin structure in response to stimuli. Additionally, unregulated promiscuous ATPase activity from all of the many chromatin remodellers present in each cell may unnecessarily deplete cellular pools of ATP or disrupt normal chromatin structure (Flaus et al., 2006, Narlikar et al., 2013). Indeed, overexpression of the Rhp26 chromatin remodeller in *Schizosacchromyces pombe* resulted in severe cellular toxicity in the absence of exogenous cellular stress. Strikingly, this phenotype was more severe when a hyperactive mutant was overexpressed, suggesting a casual correlation between ATPase hyperactivity and cellular toxicity (Wang et al., 2014a).

1.2.4 Chromatin Remodellers as Multi-Subunit Protein Complexes

Chromatin remodellers frequently manifest as large, multi-subunit protein complexes (Cairns et al., 1996, Tong et al., 1998, Xue et al., 1998, Zhang et al., 1998, Wade et al., 1999, Clapier and Cairns, 2009). As accessory subunits may contribute unique histone mark-reading domains or protein-protein interaction surfaces, it is conceivable that they may drastically influence the function of the catalytic subunit, for instance, by modulating its genomic localization or its interaction partners (Mohrmann and Verrijzer, 2005). Hence, by associating with different sets of accessory subunits, the same catalytic subunit may be a component of multiple discrete complexes, with each complex being specialized for context-specific cellular functions. Although this phenomenon has been termed combinatorial assembly, it does not mean that all permutations of subunits are

functionally active or can be purified from cells as *bona fide* protein complexes. Instead, the current data suggests that in certain tissues, canonical subunits may be replaced in chromatin remodelling complexes by related homologues (Lickert et al., 2004, Lessard et al., 2007, Wu et al., 2007, Ho et al., 2009b). Thus, combinatorial subunit assembly may be particularly applicable for factors involved in development that act on different sets of target genes according to the specific developmental stage. Indeed, a multitude of variant-PRC2 complexes have been identified beyond the canonical PRC2 complex that are integral to H3K27me3-mediated repression of developmental genes (Schwartz and Pirrotta, 2013).

Combinatorial assembly in chromatin remodelling complexes has been best characterized for the BAF complex, the mammalian orthologue of *S. cerevisiae* SWI/SNF. Indeed, BAF complexes specific to ESCs, neural progenitor cells, post-mitotic neurons and myocytes are essential for the maintenance of pluripotency, self-renewal of neural progenitors, dendrite growth, and cardiac development respectively (Lickert et al., 2004, Lessard et al., 2007, Wu et al., 2007, Ho et al., 2009b). For instance, differentiation of neural progenitor cells into post-mitotic neurons is accompanied by replacement of the BAF45a and BAF53a subunits by BAF45b, BAF45c and BAF53b. Correspondingly, the neural progenitor cell-specific subunits BAF45a and BAF53a promote cell proliferation and self-renewal (Lessard et al., 2007). In contrast, BAF53b is required in post-mitotic neurons for proper activity-dependent dendritic development as it mediates recruitment of the BAF complex to the promoters of specific target genes (Wu et al., 2007).

The capacity for small alterations in complex subunit composition to dramatically affect cellular phenotypes is illustrated by the potential for the mammalian SWI/SNF complex subunit, SNF5 (BAF47), to act as a tumour suppressor gene (Oruetxebarria et al., 2004, Vries et al., 2005, Kadoch and Crabtree, 2013). Deletion or mutation of SNF5 is observed in effectively all instances of malignant rhabdoid tumours (Versteege et al., 1998). However, re-expression of SNF5 in a malignant rhabdoid tumour cell line dramatically inhibits cell proliferation by inducing senescence. This dramatic reversal in phenotype is achieved by SNF5 mediating recruitment of the SWI/SNF complex to the $p16^{INK4a}$ promoter, resulting in expression of the tumour suppressor $p16^{INK4a}$ (Oruetxebarria et al., 2004).

Strikingly, although the loss of SNF5 in malignant rhabdoid tumour cells is associated with polyploidy and chromosomal instability, these phenotypes can be reversed by re-expression of SNF5 (Vries et al., 2005).

Curiously, cancers have evolved an alternative mechanism to inactivate SNF5 – through combinatorial subunit assembly, or more precisely, disassembly. Synovial sarcomas are characterized by the t(X;18)(p11.2;q11.2) chromosomal translocation that results in the expression of the SS18-SSX fusion protein, which competitively displaces SS18 and SNF5 from the BAF complex. This alteration in subunit composition causes aberrant localization of the mutant BAF complex at the *SOX2* locus, ultimately resulting in the transcriptional de-repression of the proto-oncogene, SOX2 (Kadoch and Crabtree, 2013).

Another ubiquitous chromatin remodelling complex that undergoes combinatorial assembly is the NuRD complex that typically contains some combination of the following subunits: the ATPases CHD3 and CHD4, histone deacetylases 1 and 2 (HDAC1 and HDAC2); methyl-CpG-binding domain 2 (MBD2) and MBD3; metastasis-associated 1 (MTA1), MTA2 and MTA3; the zinc-finger (ZNF) proteins GATAD2A (p66α) and GATAD2B (p66β); the histone-binding proteins retinoblastoma-binding protein 4 (RBBP4) and RBBP7; and the tumour suppressor deleted in oral cancer 1 (DOC1) (Tong et al., 1998, Xue et al., 1998, Zhang et al., 1998, Wade et al., 1999, Le Guezennec et al., 2006). It has been observed that the presence of MBD2 and MBD3 in the NuRD complex is mutually exclusive (Le Guezennec et al., 2006). Yet, most cell types appear to express both MBD2 and MBD3, apart from ES cells, which specifically express MBD3 (Kaji et al., 2006, Gunther et al., 2013). In ES cells, MBD3 is necessary for the assembly of a stable NuRD complex. While ES cells generally need to be cultured in the presence of the cytokine, leukaemia inhibitory factor (LIF), for the maintenance of their pluripotent phenotype, ES cells lacking MBD3 ($MBD3^{-/-}$) could be maintained in a pluripotent state despite culture in the absence of LIF. Conversely, however, $MBD3^{-/-}$ ES cells manifest defects in differentiation due to an inability to commit to developmental lineages (Kaji et al., 2006). These tissue-specific properties of MBD3 are consistent with the observation that $MBD3^{-/-}$ knockout mice experience embryonic lethality early in embryogenesis, in contrast to $MBD2^{-/-}$ knockout mice, which are viable and

fertile (Hendrich et al., 2001). At a biochemical level, while the MBD2/NuRD complex binds to methylated DNA, MBD3/NuRD is incapable of doing so (Zhang et al., 1999, Le Guezennec et al., 2006). Indeed, this observation is in accord with whole-genome studies characterizing the genomic localization of MBD2 and MBD3 by chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq). Though CpG islands – regulatory regions of genes near the TSS containing a cytosine-guanine dinucleotide sequence – were the principal binding site of both proteins, a high proportion of the CpG islands bound by MBD2 were methylated, in contrast to MBD3 that primarily bound to un-methylated genes. Moreover, MBD2-bound genes typically had low transcriptional activity, in contrast to the relatively high levels of transcription displayed by genes bound by MBD3. In fact, recruitment of MBD2-NuRD to a LacO array in cells is sufficient to mediate the conversion of open euchromatin into highly compacted heterochromatin bearing repressive marks (Gunther et al., 2013). Besides differences in genomic localization, MBD2 and MBD3 also modulate NuRD function by mediating specific protein-protein interactions. For example, the arginine methyltransferase, PRMT5, has been described to be a specific interaction partner of the MBD2/NuRD complex, via an interaction with an N-terminal region of MBD2 rich in arginine and glycine residues. Indeed, PRMT5 is recruited to some CpG islands in an MBD2- and DNA-methylation-dependent manner (Le Guezennec et al., 2006).

All three homologues, MTA1, MTA2 and MTA3, have been identified to biochemically co-purify with the NuRD complex (Xue et al., 1998, Zhang et al., 1998, Zhang et al., 1999, Fujita et al., 2003, Le Guezennec et al., 2006). However, immuno-affinity purification of the complex using an antibody specific to MTA2 failed to recover either MTA1 or MTA3, and *vice versa*, arguing against the simultaneous presence of all three homologues in the same NuRD complex (Zhang et al., 1999, Fujita et al., 2003, Yao and Yang, 2003). Transiently over-expressed Gal4-MTA1 and Gal4-MTA2 fusion proteins functioned as potent transcriptional repressors of a luciferase reporter gene. In the same assay, however, Gal4-MTA3 was a significantly less effective transcriptional repressor, suggesting that despite their sequence similarity, these homologues may mediate subtly different biological effects (Yao and Yang, 2003). This is exemplified in the context of breast cancer cells, where expression of MTA3 requires oestrogen receptor positivity;

correspondingly, therefore, incorporation of MTA3 into the NuRD complex is an oestrogen-dependent process. MTA3 – in contrast to MTA1, for example – mediates transcriptional repression of the transcriptional factor, Snail, probably in a direct manner as MTA3 can be detected on the *Snail* gene promoter by ChIP. As Snail inhibits expression of E-cadherin, an important cell-adhesion molecule specific to epithelial cells, it is unsurprising that the absence of MTA3 in breast cancer cells is associated with increased levels of Snail and down-regulation of E-cadherin. The loss of expression of E-cadherin is believed to be a key contributor to an epithelial-to-mesenchymal cell transition, a process that is associated with the tumour invasion (Fujita et al., 2003). While the precise roles of the MTA proteins in the NuRD complex remain to be comprehensively elucidated, in an *in vitro* system, MTA2 facilitates the assembly of an enzymatically-active histone deacetylase complex – a core complex, containing HDAC1, HDAC2, RBBP4 and RBBP7, displayed considerably more histone deacetylase activity in the presence of MTA2 than in its absence (Zhang et al., 1999). Nevertheless, it remains unclear whether its homologues MTA1 and MTA3 are similarly capable in this regard.

Collectively, as illustrated by these observations, the ease with which the subunit composition of chromatin remodelling complexes can be manipulated to dramatically affect cellular phenotypes emphasizes the flexibility and power of combinatorial subunit assembly as a regulatory mechanism.

1.3 CHD1

1.3.1 Domain Architecture of CHD1

CHD1 is a highly conserved chromatin remodeller. Indeed, the domain architecture of *S. cerevisiae* Chd1 is largely comparable to that of human CHD1 (Marfella and Imbalzano, 2007). The main features of CHD1 are a pair of N-terminal tandem chromodomains, a centrally located ATPase domain, and a C-terminal DNA-binding domain that is structurally homologous to the SANT and SLIDE domains of ISWI-type chromatin remodellers (Figure 1.10) (Clapier and Cairns, 2009, Ryan et al., 2011, Sharma et al., 2011).

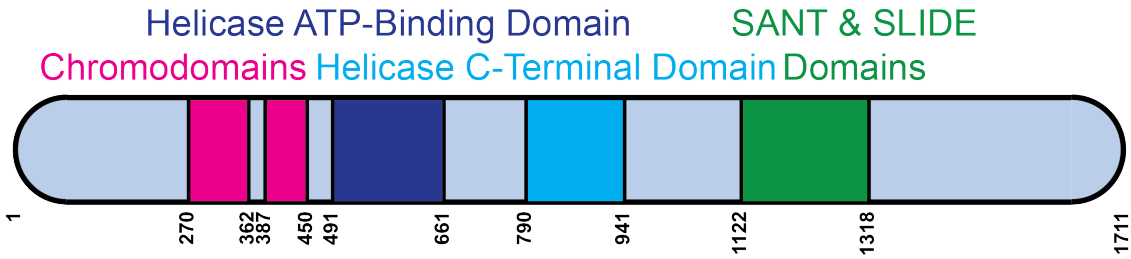


Figure 1.10 – Domain Architecture of Mouse CHD1
This schematic of mouse CHD1 illustrates its annotated domains.

Chromodomains typically function as histone methyl-lysine recognition modules (as discussed in subsection **1.1.4**). While the chromodomains of human CHD1 behave predictably and specifically recognize the active H3K4me3 modification, the chromodomains of yeast Chd1 do not bind to (mono-, di-, or tri-) methylated H3K4 (Flanagan et al., 2005, Sims et al., 2005). Nonetheless, the chromodomains of CHD1 do not recognize the methylated histone tail in the conventional manner since a helix-turn-helix motif connects the two chromodomains and positions them contiguously, forming a single, continuous surface for binding histone tails (Figure 1.11). Rather than the conventional three-residue aromatic cage for methyl-lysine recognition, the CHD1 chromodomains use only two tryptophan residues (64 and 67) – one contributed by each chromodomain. Thus, the methylammonium of lysine 4 is not encapsulated by a π-electron cage, as is typical for most chromodomain-methylated histone tail interactions. Instead, the interaction needs to be further stabilized by a cation-π interaction between tryptophan 67 and arginine 2 of the histone tail (Flanagan et al., 2005). Moreover, the surfaces employed by the tandem CHD1 chromodomains to interact with the H3K4me3 tail differ from that of canonical chromodomains, since many of the usual interaction interfaces are blocked in CHD1 by unique structural inserts (Patel and Wang, 2013). Therefore, structural studies clearly demonstrate that despite not being organized as canonical chromodomains, the tandem chromodomains of CHD1 nevertheless act as a *bona fide* H3K4me3-recognition module.

In addition to recognizing the H3K4me3 histone modification, the tandem chromodomains of CHD1 probably also negative regulate the ATPase domain. Structural studies of budding yeast Chd1 reveal that the chromodomains are wedged between the two lobes of the ATPase domain, occupying its DNA-binding surface and also locking it into an inactive ‘open’ conformation (as previously

discussed in further detail in subsection **1.2.3**) (Hauk et al., 2010). It is presumed that mammalian CHD1 proteins are similarly subjected to negative regulation by modular allostery of their chromodomains, though this assumption awaits formal validation. Moreover, it remains unclear how binding of the chromodomains to histone tails influence auto-inhibition of the ATPase domain. One conceptually attractive model is that recognition of the active H3K4me3 modification by the chromodomains relieves auto-inhibition, thereby allowing the chromodomains to restrict catalytic activity of the CHD1 chromatin remodeller to appropriate – actively transcribed – sites in the genome.

Figure 1.11 – Recognition of the H3K4me3 Mark by the Tandem Chromodomains of Human CHD1
The tandem chromodomains of human CHD1 act cooperatively to interact with the H3K4me3 mark by forming a single, continuous interaction surface. Instead of a conventional three-sided aromatic cage for methyl-lysine recognition, the CHD1 chromodomains use only two tryptophan residues (64 and 67) – one contributed by each chromodomain. The interaction is further stabilized by a cation-π interaction between tryptophan 67 and arginine 2 of the histone tail. The image was taken from (Patel and Wang, 2013).

The C-terminal DNA-binding domain of *S. cerevisiae* Chd1 bears structural homology to the SANT and SLIDE domains of ISWI-type chromatin remodellers despite sharing little sequence homology (Ryan et al., 2011, Sharma et al., 2011). The SANT and SLIDE domains are necessary and sufficient for binding to DNA and nucleosomes (Ryan et al., 2011). Curiously, however, though dispensable for Chd1 nucleosome sliding activity *in vitro*, the DNA-binding domain nonetheless dictates the directionality of sliding, which for Chd1 is to centre end-positioned nucleosomes (McKnight et al., 2011).

In contrast to multi-subunit chromatin remodelling complexes, budding yeast and *Drosophila* CHD1 has been described to function in isolation without any accessory subunits (Tran et al., 2000, Lusser et al., 2005, Ehrensberger and Kornberg, 2011). Correspondingly, the structural and biochemical evidence described above clearly indicate that CHD1 intrinsically possesses all the necessary components to function as an autonomous chromatin remodeller (Flanagan et al., 2005, Hauk et al., 2010, Ryan et al., 2011). Yet, it is unclear how closely mammalian CHD1

proteins recapitulate the properties of their budding yeast homologue. Therefore, further biochemical characterization is required to elucidate the mechanistic basis for the developmental roles ascribed to CHD1 in higher eukaryotes.

1.3.2 Functions of CHD1

Biochemically, budding yeast Chd1 demonstrates nucleosome sliding activity *in vitro*. Chd1 shifts end-positioned nucleosomes to the centre of the DNA fragment, in a manner similar to ISWI-type enzymes (Stockdale et al., 2006). In addition, *Drosophila* CHD1 acts in concert with the histone chaperone, NAP1, to deposit histone octamers on DNA and to generate evenly spaced nucleosomal arrays in an ATP-dependent manner (Lusser et al., 2005). Due to the conservation of CHD1 through evolution, it is likely that the CHD1 homologues of higher eukaryotes possess comparable biochemical activities.

It was initially observed that CHD1 localizes to active regions of *Drosophila* polytene chromosomes, including puff and interband regions (Stokes et al., 1996). Genome-wide investigation of CHD1 localization at a molecular level by chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) in MEFs has revealed that CHD1 is enriched at promoters. Moreover, there is a high correlation between promoter occupancy of CHD1 and RNAPII, strongly suggesting that CHD1 is recruited to the promoters of transcriptionally active genes (Skene et al., 2014). Further suggesting a role in transcription, proteomic and yeast two-hybrid screens indicate that Chd1 interacts with transcription elongation factors in budding yeast (Krogan et al., 2002, Simic et al., 2003).

CHD1 has been implicated to influence various stages of the transcription cycle. In a fully defined *in vitro* system, *S. cerevisiae* Chd1 ejects the promoter nucleosome at the repressed *PHO5* gene locus in a gene activator-dependent manner. Confirming a role in promoting transcriptional initiation, *PHO5* gene expression was repressed *in vivo* in a *chd1Δ, isw1Δ* strain, lacking both the Chd1 and Isw1 chromatin remodellers (Ehrensberger and Kornberg, 2011). A role as a termination factor was ascribed to Hrp1, a CHD1 homologue in *S. pombe*, by genetic screening.

Strikingly, this role in transcription termination is evolutionarily conserved since Chd1 also regulates termination at the *CYC1* and *ASC1* loci in *S. cerevisiae*, a distant yeast species, by modulating chromatin structure at the 3' end of those genes, though the precise mechanism remains undefined (Alen et al., 2002). In contrast, CHD1 does not appear to facilitate pre-mRNA splicing through changes in chromatin structure. CHD1 stably associates with the SF3a subcomplex of the spliceosome, and directly promotes recruitment of the spliceosome to chromatin by physically bridging an interaction with histone tails bearing the H3K4me3 modification (Sims et al., 2007).

In budding yeast, Chd1 adopts a dominant role, albeit assisted by Isw1, in limiting histone exchange within the open reading frame of genes, especially towards their 3' ends (Smolle et al., 2012). This is probably a conserved role for CHD1, since it also limits nucleosome turnover within the gene body in MEFs (Skene et al., 2014). Considering its ability to catalyze nucleosome assembly *in vitro*, CHD1 probably promotes the nucleosome reassembly in the wake of transcription elongation (Lusser et al., 2005, Skene et al., 2014). Supporting a crucial role in maintaining global chromatin structure, increased cryptic transcription has been reported in a *chd1Δ* budding yeast strain (Smolle et al., 2012). One notable difference in the regulation of nucleosome dynamics by different CHD1 orthologues has, nevertheless, been noted – in MEFs (and presumably in most mammalian cells) but not in budding yeast, nucleosome turnover at the vast majority of promoters is controlled by CHD1 (Skene et al., 2014).

In addition to regulating nucleosome occupancy, CHD1 also influences nucleosome positioning at a genome-wide level in several species. Although deletion of the *chd1* gene by itself is insufficient for genome-wide changes in nucleosome positioning in *S. cerevisiae*, the combined absence of the CHD1 and ISW1 chromatin remodellers disrupts nucleosome positioning in the gene body, though without affecting the +1 promoter nucleosome (Gkikopoulos et al., 2011). Likewise, the *S. pombe* CHD1 homologues, Hrp1 and Hrp3, are required to maintain proper nucleosome positioning in the gene body (Pointner et al., 2012, Shim et al., 2012). However, evidence implicating CHD1 as a genome-wide determinant of nucleosome positioning in mammalian cells has not been

forthcoming thus far – while depletion of CHD1 in MEFs reduces global nucleosome occupancy, nucleosome positioning is unaffected (Skene et al., 2014). Nevertheless, it is entirely plausible that akin to the situation in yeast, functional redundancy enables higher eukaryotes to compensate for the absence of CHD1 by relying on other chromatin remodellers to ensure proper nucleosome positioning.

Given its importance for the regulation of transcription and the maintenance of chromatin organization, it would be unsurprising if the absence of CHD1 from cells were to manifest as striking cellular phenotypes. Indeed, mouse ESCs depleted of CHD1 display reduced self-renewal ability and compromised pluripotency, reflecting an inability to give rise to primitive endoderm (Gaspar-Maia et al., 2009). Notably, even mutations that are ostensibly minor can compromise normal differentiation – deletion of the first hundred amino acids from CHD1, which does not affect any of the annotated domains, is sufficient to skew *in vitro* differentiation of mouse ESCs in favour of the neuroectodermal lineage (Piatti et al., 2015). Additionally, depletion of CHD1 from mESCs causes an increase in heterochromatic foci marked by the H3K9me3 modification, suggesting that CHD1 is an important regulator of the open chromatin state characteristic of pluripotency (Gaspar-Maia et al., 2009). Nevertheless, the role of CHD1 in pluripotency and development can still only be breached in descriptive terms, since mechanistic explanations for these fascinating observations remain elusive. Thus, the possibility of pluripotency-specific CHD1 complex subunits was investigated in this study (Chapter 3) by purification and mass spectrometry analysis of the endogenous CHD1 complex from mouse ESCs.

1.4 CHD5

1.4.1 CHD3, CHD4 and CHD5 are Homologues

CHD5 is a homologue of CHD3 and CHD4, sharing 68% and 72% sequence identity with each protein respectively. Yet, one notable difference between them is their pattern of expression – CHD3 and CHD4 are ubiquitously expressed, whereas CHD5 is restricted to the nervous system and testis (Bergs et al., 2014, Zhuang et al., 2014). Indeed, the expression of CHD5 is mutually exclusive to that of CHD3

and CHD4 during spermatogenesis. CHD5 expression is undetectable in pre-meiotic spermatogonia and spermatocytes, where CHD3 and CHD4 are highly abundant. In contrast, CHD5 expression reaches its zenith in post-meiotic late round spermatids, where CHD3 and CHD4 are restricted to low expression levels (Bergs et al., 2014).

CHD3 and CHD4 are catalytic subunits of the NuRD complex. The NuRD complex may include the following components: histone deacetylases 1 and 2 (HDAC1 and HDAC2); methyl-CpG-binding domain 2 (MBD2) and MBD3; metastasis-associated 1 (MTA1), MTA2 and MTA3; the zinc-finger (ZNF) proteins GATAD2A (p66α) and GATAD2B (p66β); the histone-binding proteins retinoblastoma-binding protein 4 (RBBP4) and RBBP7; and the tumour suppressor deleted in oral cancer 1 (DOC1) (Tong et al., 1998, Xue et al., 1998, Zhang et al., 1998, Wade et al., 1999, Le Guezennec et al., 2006). The NuRD complex is unique in its possession of two distinct catalytic activities – nucleosome remodelling and histone deacetylation. Since histone acetylation is a mark of transcriptionally active genes, it is unsurprising that a major function of the NuRD complex is transcriptional repression (Lai and Wade, 2011).

It has been observed that CHD5 co-immunoprecipitates with subunits of the NuRD complex, including HDAC2, RBBP7 and MTA3. Hence, it was speculated that CHD5 is a component of the NuRD complex in neuronal tissues (Potts et al., 2011). Nonetheless, this hypothesis has not been formally tested and it remains unclear whether CHD5 is stably associated in the NuRD complex. Due to the homology between these chromatin remodellers, CHD5 could conceivably either replace CHD3 and CHD4 in the NuRD complex, or be incorporated along with them. Further biochemical characterization is, consequently, integral to acquiring a detailed understanding of CHD5 and the complex in which it carries out its cellular functions. Moreover, it is not inconceivable that the NuRD complex displays combinatorial subunit assembly with regards to its catalytic subunit, since this phenomenon has already been documented for both its MBD and MTA subunits (as previously discussed in subsection **1.2.4**) (Fujita et al., 2003, Le Guezennec et al., 2006).

1.4.2 Domain Architecture of CHD5

A member of the Mi2 subfamily of chromatin remodellers, CHD5 contains four histone modification-reading domains in its N-terminus – a pair of tandem PHD domains followed by tandem chromodomains (Figure 1.12). The tandem PHD domains of CHD5 specifically recognize unmodified histone H3 tails (Paul et al., 2013). Structural studies of DPF3b (also known as BAF45c; a subunit of the BAF complex in post-mitotic neurons) and the histone acetyltransferase MOZ reveal that their tandem PHD domains are positioned contiguously to form a single, continuous interaction surface to recognize histone H3 tails acetylated at lysine 14 (Zeng et al., 2010, Qiu et al., 2012). In contrast, each PHD domain of CHD5 is capable of independently recognizing an unmodified histone H3 tail. Nevertheless, they work cooperatively to interact simultaneously with two histone H3 molecules; through this mode of bivalent interaction, the tandem PHD domains benefit from four- to eleven-fold greater affinity for unmodified H3 tails compared to either individual domain (Oliver et al., 2012). Indeed, the tandem PHD domains of the homologous CHD4 protein behave in a comparable manner to interact with two histone H3 tails marked by the H3K9me3 modification. It was also suggested that the tandem CHD4 PHD domains bind to the two H3 tails of a single nucleosome. This hypothesis arises from the inference (based on nuclear magnetic resonance (NMR) data indicating that the linker connecting the CHD4 PHD domains is largely unstructured) that the distance separating the two PHD domains is approximately 70Å, which coincides with the distance between the two H3 tails of a single nucleosome (Musselman et al., 2012b).

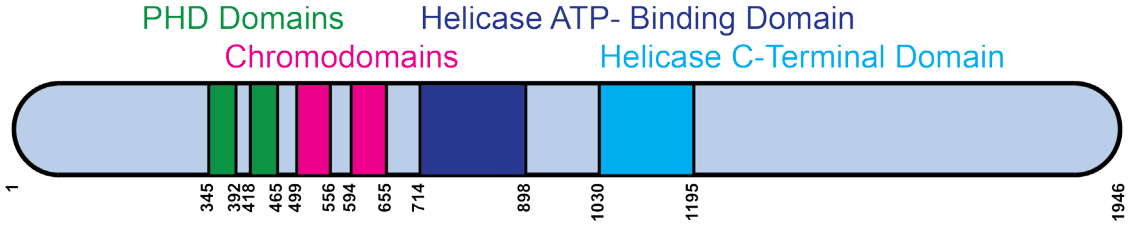


Figure 1.12 – Domain Architecture of Mouse CHD5
This is a schematic representation of mouse CHD5, showing its annotated domain features.

The tandem chromodomains also mediate histone mark recognition, specifically the H3K27me3 modification (Egan et al., 2013). Yet, in the absence of an atomic

structure, it remains unclear whether the CHD5 tandem chromodomains are independently capable of recognizing the H3K27me3 mark, or whether they form a single unified histone interaction surface comparable to that of the CHD1 chromodomains (Flanagan et al., 2005).

Considered as a whole, the N-terminal histone mark-binding region of CHD5 displays highest affinity for the H3K27me3 modification, though it also interacts specifically with unmodified H3 tails (Egan et al., 2013). The dominance exerted by the chromodomains in ignoring the preference of the PHD domains for unmodified H3 tails is probably a simple reflection of their respective interaction affinities. Indeed, the dissociation constant (K_d) governing the interaction between the chromodomains and an H3K27me3 peptide is 1nM, whereas the PHD domains bind unmodified H3 tail peptides with a K_d of 9μM (Oliver et al., 2012, Egan et al., 2013).

The PHD and chromodomains of CHD4 are essential for the transcriptional repression of a model gene, *mb-1* (*Cd79a*), in B-lymphocytes. As mutation of these histone mark-reading domains compromises the recruitment of CHD4 to the *mb-1* promoter, it is likely the PHD and chromodomains contribute to CHD4-mediated transcriptional repression partly by regulating the genomic localization of CHD4 (Musselman et al., 2012b, Ramirez et al., 2012). However, the tandem PHD and chromodomains probably also influence the ATPase motor through allosteric regulation (Morra et al., 2012). Given the similarities between CHD4 and CHD5, it is likely that the tandem PHD and chromodomains also regulate the recruitment of CHD5 to chromatin. Yet, limited correlation was identified between the genomic localization of CHD5 and the H3K27me3 modification recognized by its tandem chromodomains (Egan et al., 2013). Thus, further work is required to understand the interplay between the PHD domains and chromodomains; specifically, it is unclear whether they function entirely independently of one another, or whether they bind histone tails cooperatively. Indeed, it must be clarified whether the histone mark recognition domains in CHD5 form an integrated interaction module, such as that of the PHD-bromodomain of TRIM24, which interacts with a histone tail that is unmodified at lysine 4, but acetylated at lysine 23 (Tsai et al., 2010). Ultimately, clarification of the structural basis for the interaction between CHD5 and

histone tails will be necessary to fully comprehend the determinants controlling the recruitment of CHD5 to chromatin and to reconcile biochemical and genome-wide ChIP-seq data.

The ATPase domain and C-terminus of CHD5 have not been characterized biochemically. Indeed, a DNA-binding domain similar to the SANT and SLIDE domains of CHD1 and ISWI-type chromatin remodellers has not been identified in CHD5. Yet, pure recombinant CHD5 is able to interact with nucleosomes, suggesting that similar to other SWI2/SNF2 chromatin remodelling ATPases, the core ATPase motor of CHD5 is capable of binding DNA and thereby, nucleosomes (Durr et al., 2005, Mueller-Planitz et al., 2013, Quan and Yusufzai, 2014).

1.4.3 Functions of CHD5

Recombinant CHD5 is capable of increasing the accessibility of chromatinized templates and mononucleosomes to restriction enzymes *in vitro* (Quan and Yusufzai, 2014). However, this chromatin remodelling activity appears to be distinct from either the nucleosome sliding typical of most chromatin remodellers or the nucleosome ejection catalysed by the RSC complex (Lorch et al., 1999, Wang and Zhang, 2001, Stockdale et al., 2006). CHD5-remodelled chromatin was subjected to micrococcal nuclease digestion followed by glycerol gradient sedimentation, which separates according to molecular mass. Although the CHD5-remodelled sample contained mono- and di-nucleosomes that co-sedimented with canonical mono- and di-nucleosomes, the nucleosome fractions were associated with DNA of the expected length, along with shorter DNA fragments. This observation is inconsistent with nucleosome sliding, since repositioned nucleosomes would nonetheless protect the same length of DNA. Instead, CHD5-remodelling increases the accessibility of endonucleases to nucleosomal DNA that is usually protected by histone-DNA interactions. This was conceptualized to represent CHD5 stably unwrapping DNA from the histone octamer at an internal nucleosomal site, approximately 40-50bp from the entry or exit point, in an ATP-dependent manner. This form of chromatin remodelling appears to be unique to CHD5, as nucleosome unwrapping activity is not observed even with its close homologue, CHD4. Overall,

in terms of chromatin remodelling, CHD5 acts with low processivity and does not appear to affect nucleosome positioning (Quan and Yusufzai, 2014).

The effect of CHD5 on chromatin organization *in vivo* is unknown. However, given the similarity between CHD5 and CHD4, and the link between both proteins and the NuRD complex, it is possible that they function similarly (Potts et al., 2011). Genome-wide DNase I hypersensitivity mapping has revealed that inhibition of CHD4 function by inducing over-expression of an 'ATPase-dead', dominant-negative mutant, which presumably outcompetes endogenous CHD4 for genomic binding sites, increases chromatin accessibility at a large number of sites (4125 sites) across the genome. In contrast, considerably fewer DNase I hypersensitive sites are lost (242 sites) in response to antagonism of CHD4. Thus, this data indicates a global role for CHD4 at maintaining a closed chromatin structure (Morris et al., 2014). Similarly, in *Drosophila*, the NuRD complex acts at its target sites primarily by increasing nucleosome density; notably, these sites generally possess sequence properties that discourage nucleosome occupancy (Moshkin et al., 2012). Collectively, these observations are entirely consistent with the prevailing paradigm of CHD4 as a transcriptional repressor (Lai and Wade, 2011).

Extrapolating from the role of CHD4 as a transcriptional repressor, it might be expected for CHD5 to adopt a similar function. However, there is currently only limited data to support this notion. For instance, CHD5 represses the *WEE1* gene in human cells (Quan et al., 2014). However, gene expression microarray analysis of neural progenitor cells, derived either from control mESCs or from mESCs depleted of CHD5, identified 512 genes that were preferentially activated during neurogenesis in control but not in CHD5-depleted cells. In contrast, only 119 genes were specifically activated in CHD5-depleted compared to control cells. Therefore, this data suggests that CHD5 acts more frequently to activate rather than repress transcription (Egan et al., 2013). Yet, a notable caveat is that it is impossible to distinguish genes that are directly regulated by CHD5 from those whose expression is only indirectly influenced by CHD5. Moreover, as this result reflects the function of CHD5 at a specific stage in neuronal development, it may not be indicative of the manner in which CHD5 regulates transcription under normal circumstances.

As suggested by its restricted expression pattern, CHD5 is involved in developmental processes. In the context of neural development, CHD5 upregulates neuronal genes, especially those encoding factors involved in late-stage neural development. Furthermore, depletion of CHD5 from the developing mouse neocortex prematurely curtails neuronal differentiation, as reflected in a surfeit of immature neural progenitor cells (Egan et al., 2013). Yet, *CHD5* knockout mice display largely unremarkable neurological phenotypes, exhibiting only mild neuro-behavioural defects (Zhuang et al., 2014). Considered together, it is probable that neuronal differentiation is regulated by multiple important, yet redundant mechanisms, of which CHD5 is but one.

The dominant phenotype of *CHD5* knockout mice is male infertility (Zhuang et al., 2014). As befitting a regulator of sperm development, expression of CHD5 in the testis is temporally regulated relative to spermatogenesis, reaching a peak in post-meiotic late round spermatids despite being virtually absent in pre-meiotic spermatogonia and spermatocytes (Bergs et al., 2014). Loss of CHD5 impairs spermiogeneis due to compromised chromatin integrity and compaction in sperm, which probably arise due to deregulation of the histone to protamine transition. Compared to wild-type mice, differentiated spermatids in CHD5-deficient mice retained higher levels of core histones for longer in spermiogenesis; moreover, CHD5 deficiency was also associated with perturbations in the expression of protamines and important regulators of the histone to protamine transition, such as Tnp1 and Tnp2. CHD5-deficient spermatids also exhibited reduced levels of histone H4 acetylation. Histone H4 hyperacetylation is a key molecular event for protamine incorporation as it induces histone eviction. This is probably achieved by a direct association between the acetylated histone tails and the testis-specific, bromodomain-containing protein, Brdt; it is further speculated that Brdt, in turn, recruits chromatin remodellers to eject histone octamers from nucleosomes. It is clear that CHD5 influences many processes integral to the histone to protamine replacement process; however, the roles of CHD5 in this process have yet to be precisely elucidated – it is unknown whether CHD5 acts indirectly by modulating gene expression or directly influences chromatin structure (Li et al., 2014).

CHD5 is not only associated with physiological development, but also with pathology – CHD5 has been identified to be the main tumour suppressor gene compromised by the *1p36* chromosomal deletion that has an incidence rate of 70-80% in high-risk neuroblastomas (Bagchi et al., 2007, Fujita et al., 2008, Okawa et al., 2008). In this setting, the absence of CHD5 promotes cellular proliferation and inhibits further neuronal differentiation of malignant cells. Although the tumour-suppressive function of CHD5 cannot be defined in mechanistic terms, it is dependent on the tandem PHD domains of CHD5, suggesting that proper genomic localization of this chromatin remodeller is essential for this process (Paul et al., 2013).

While it is clear that CHD5 is functionally distinct from its homologues, CHD3 and CHD4, it is unclear how similar the CHD5 complex composition is to that of either CHD3 or CHD4. However, given the relatively high sequence identity shared between CHD3, CHD4 and CHD5, it was predicted that the tissue-specific functions of CHD5 could be attributed to distinct complex subunits that either influenced the catalytic activity of CHD5 or regulated its genomic localization. Thus, the native CHD5 complex was purified from mouse brains and its subunit composition analysed by mass spectrometry, as described in Chapter 4.

1.5 SMARCAD1

1.5.1 Domain Architecture of SMARCAD1 – Comparisons with the Related SWR1 and INO80 Chromatin Remodellers

Largely uncharacterized, SMARCAD1 is classified by multi-sequence alignment as an Swr1-like chromatin remodeller due to its split-ATPase domain organization (Figure 1.13) (Flaus et al., 2006). Due to a dearth of biochemical characterization of the SMARCAD1 protein, it is necessary to extrapolate from our understanding of the well-characterized SWR1 and INO80 chromatin remodelling complexes to obtain just the vaguest sense of the intrinsic capabilities of SMARCAD1. Yet, it must be emphasized that beyond a long insertion dividing their ATPase domains, SMARCAD1, Swr1 and Ino80 are not overly similar proteins – most apparently, the SWR1 and INO80 complex possess more than 15 subunits, while SMARCAD1

forms a complex with only KAP1 (Rowbotham et al., 2011, Nguyen et al., 2013, Tosi et al., 2013).

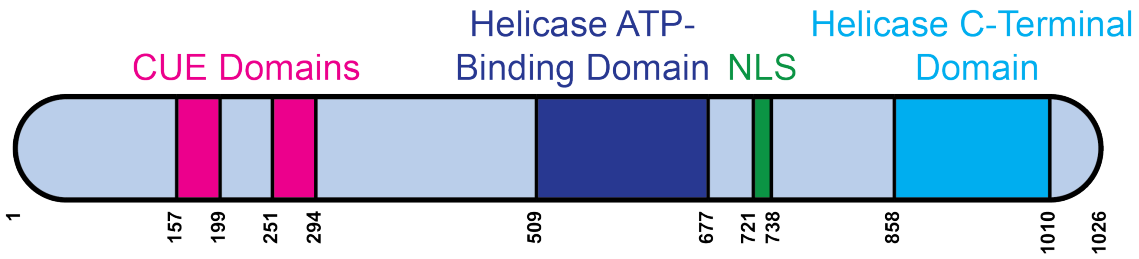


Figure 1.13 – Domain Architecture of Human SMARCAD1
This schematic of human SMARCAD1 illustrates its domain features. The long insertion between the ATPase domains characterizes SMARCAD1 as an Swr1-like remodeller. The abbreviation, NLS, refers to a nuclear localization signal.

The defining shared characteristic of Swr1 and Ino80 is the long insertion dividing their ATPase domains. This insertion mediates an interaction with a hetero-oligomer of Rvb1 and Rvb2, ATPases of the AAA+ family that are related to RuvB – a bacterial single-stranded DNA helicase that promotes strand migration for the resolution of Holliday junction structures that arise during DNA repair by homologous recombination. It is unclear whether human Rvb1/Rvb2 heterodimers possess helicase activity (Gerhold and Gasser, 2014). Nevertheless, mediated by the split insert domains of Swr1 and Ino80, the SWR1 complex incorporates a heterohexameric Rvb1/Rvb2 ring, while the INO80 complex stably associates with an Rvb1/Rvb2 heterododecamer, arranged in the form of two stacked hexameric rings (Nguyen et al., 2013, Tosi et al., 2013). Proteomic analysis, however, showed that neither Rvb1 nor Rvb2 is a specific interaction partner of SMARCAD1 (Rowbotham et al., 2011). Though the ATPase domain insertion in SMARCAD1 is not directly comparable to those of Swr1 or Ino80, it may nevertheless, function an interaction surface for proteins other than Rvb1 and Rvb2. Also, without an HSA domain, SMARCAD1 does not interact with actin or actin-related proteins, in contrast to Swr1 and Ino80 (Rowbotham et al., 2011, Gerhold and Gasser, 2014).

The structures of the SWR1 and INO80 chromatin remodelling complexes have recently been defined by cryo-EM (Nguyen et al., 2013, Tosi et al., 2013). Significantly, these studies have revealed that SWR1 and INO80 interact with nucleosomes in idiosyncratic manners that bear no resemblance to each other, or

to the way in which the RSC complex engulfs a nucleosome in its central cavity (Figure 1.14) (Chaban et al., 2008, Nguyen et al., 2013, Tosi et al., 2013). When a nucleosome is bound to the compact SWR1 complex, the nucleosome is peripherally located and occupies an interaction surface formed mainly by the catalytic Swr1 subunit, but also with a contribution from the Rvb1/Rvb2 heterohexamer. Thus, in this structure, the majority of contacts between the SWR1 complex and the nucleosome involve the Swr1 subunit (Nguyen et al., 2013). Biochemical data demonstrating the preference of the SWR1 complex for nucleosomes with long regions of extra-nucleosomal DNA has, however, revealed that a DNA-binding basic region in the N-terminus of the Swc2 subunit is the main determinant of the interaction (Ranjan et al., 2013). This discrepancy has prompted speculation that the SWR1 complex undergoes a conformation change to perform histone dimer exchange (Nguyen et al., 2013, Gerhold and Gasser, 2014). In contrast, the INO80 complex resembles an elongated embryo, which can be sub-divided into head, neck, body and foot components. To bind a nucleosome, the INO80 grasps it with its head and foot modules, forming multiple nucleosome-remodeller contacts with all INO80 modules (Tosi et al., 2013). The contrasting styles employed by the related SWR1 and INO80 complexes to interact with nucleosomes preclude an accurate, *a priori* prediction of the means by which SMARCAD1 may bind nucleosomes.

Figure 1.14 – Nucleosome Binding by the INO80 & SWR1 Chromatin Remodellers

A. The INO80 complex forms an elongated embryo structure, which can be sub-divided into head, neck, body and foot components. INO80 binds nucleosomes by grasping it with its head and foot modules. The image was taken from (Tosi et al., 2013).

B. The nucleosome binds to a peripheral location on the compact SWR1 complex. The interaction interface on SWR1 is formed mainly by the catalytic Swr1 subunit, but with a contribution from the Rvb1/Rvb2 heterohexamer. The blue and yellow arrows illustrate the main conformational changes that accompany nucleosome binding to the SWR1 complex. The image was adapted from (Nguyen et al., 2013).

DNA-binding domains are crucial elements of chromatin remodellers. The importance of the SANT and SLIDE domains to ISWI-like (as part of a larger HSS motif) and Chd1-like chromatin remodellers has been previously described (in subsection **1.2.3**) (Grune et al., 2003, Ryan et al., 2011). The SWR1 complex displays a clear *in vivo* preference for binding to promoter nucleosomes flanked by

at least 50bp of linker DNA. The basis for this can be directly attributed to the biochemical properties of SWR1, which forms contacts with both the extranucleosomal DNA and the adjacent nucleosome core particle. The DNA-binding domain on the Swc2 subunit is the dominant contributor to SWR1-nucleosome binding, though the catalytic Swr1 subunit is also capable of mediating an interaction with nucleosomes, albeit to a lesser extent (Ranjan et al., 2013). Apart from the ATPase domain, SMARCAD1 does not appear to possess any motifs that will confer DNA-binding ability. Therefore, SMARCAD1 either relies exclusively on the DNA-binding ability of its ATPase motor, or possesses a hitherto un-annotated DNA-binding domain, or has evolved a novel mechanism for stabilizing its presumed interaction with nucleosomes.

Of especial relevance to the work in this thesis, a unique feature of SMARCAD1 is a pair of N-terminal tandem CUE (Coupling of ubiquitin conjugation to endoplasmic reticulum degradation) domains. CUE domains are ubiquitin-binding domains (and will be discussed in greater detail in section **1.7**). The functions of the SMARCAD1 CUE domains are unknown; however, the CUE domain of Fun30, the *S. cerevisiae* homologue of SMARCAD1, apparently does not recognize ubiquitylated histone tails (Awad et al., 2010). As the only domains annotated in SMARCAD1 besides the conserved ATPase, the hypothesis that the CUE domains regulate SMARCAD1 function, perhaps by mediating protein-protein interactions, was investigated as described in Chapter 7.

1.5.2 Functions of SMARCAD1

The biochemical properties of SMARCAD1 remain unstudied. Its budding yeast homologue, Fun30, exhibits a limited ability to slide nucleosomes; however, this is exceeded by its ability to mediate histone H2A/H2B dimer exchange (Awad et al., 2010). This activity is clearly reminiscent of the SWR1 and INO80 chromatin remodelling complexes – SWR1 incorporates free variant H2AZ/H2B dimers into nucleosomes that are originally composed of conventional H2A/H2B dimers, whereas INO80 catalyses the reverse reaction (Mizuguchi et al., 2004, Papamichos-Chronakis et al., 2011). Histone dimer exchange by SWR1 is the net

effect of at least two synergistic processes – the catalytic Swr1 subunit promotes the incorporation of H2AZ into nucleosomes, while the Swc2 subunit binds to H2AZ and acts as a lock, preventing its removal from nucleosomes (Watanabe et al., 2013). It is suggested that SWR1 unwraps nucleosomal DNA from its entry/exit sites, thereby exposing the DNA-binding surface of the H2A/H2B dimer, and facilitating the intrinsic tendency of the histone octamer to dissociate into its constituent parts, namely H2A-H2B dimers and a $(H3\text{-}H4)_2$ tetramer. With the histone octamer being in dynamic equilibrium, an Htz1-H2B dimer, brought into close proximity by the SWR1 complex, can compete for the binding site vacated by the H2A-H2B dimer. The structural incompatibility induced by the presence of two heterotypic dimers should subsequently promote SWR1-mediated exchange of the second H2A-H2B dimer (Mizuguchi et al., 2004). Histone dimer exchange appears to be conserved function for many chromatin remodellers classed as Swr1-like; consequently, SMARCAD1 may similarly possess histone dimer exchange activity.

Depletion of SMARCAD1 from human cells resulted in a global increase in acetylation of histones H3 and H4 and in a decrease in the repressive H3K9me3 mark, which is associated with constitutive heterochromatin. As SMARCAD1 co-localized with sites of DNA replication, it was suggested that SMARCAD1 functions to maintain constitutive heterochromatin, for example at pericentric repeats, through replication (Rowbotham et al., 2011). It is unclear whether SMARCAD1 achieves these functions by acting directly at the level of nucleosomes and chromatin structure, or indirectly through the regulation of gene expression. If it were the former, it is conceptually plausible that SMARCAD1 removes acetylated histones from replicated chromatin by promoting histone turnover through dimer exchange.

The link between SMARCAD1 and heterochromatin might well be attributed to KAP1 (KRAB-associated protein 1), with which most cellular SMARCAD1 forms a stoichiometric complex in human cells (Rowbotham et al., 2011). (The Krüppel-associated box (KRAB) domain is the defining feature of a family of ZNF transcription factors (Iyengar and Farnham, 2011).) By virtue of its HP1-box, the transcriptional co-repressor, KAP1 is capable of interacting with HP1 and thereby, indirectly with constitutive heterochromatin (Ryan et al., 1999). Consequently, it is

possible that KAP1 modulates the genomic localization of SMARCAD1 by recruiting it to constitutive heterochromatin, though such a hypothesis awaits empirical validation.

Intriguingly, two independent genetic screens in yeast have implicated the SMARCAD1 homologue, Fun30, as a factor that promotes DNA end resection (Chen et al., 2012, Costelloe et al., 2012). End resection – the conversion the ends of DNA double-strand breaks (DSBs) into 3' single-stranded DNA overhangs by 5'→3' exonuclease cleavage – is a requisite for the repair of DSBs by the error-free homologous recombination pathway (Jasin and Rothstein, 2013). Fun30 is believed to directly regulate end resection as it is recruited to DSBs and spreads away bidirectionally from the DNA lesion over time (Chen et al., 2012, Costelloe et al., 2012). Moreover, Fun30 interacts with canonical mediators of end resection in yeast, including the Exo1 exonuclease and RPA, a single-stranded DNA-binding protein. The mechanism by which Fun30 promotes end resection is unknown; however, it is unlikely to be achieved by regulating local nucleosome repositioning, as the absence of Fun30 fails to significantly alter nucleosome positions at DSBs (Chen et al., 2012). The observations have been extended to SMARCAD1, which is recruited to DSBs induced by laser micro-irradiation with similar kinetics to the Exo1. Additionally, in response to ionizing radiation, human fibroblasts depleted of SMARCAD1 form fewer RPA foci, which arise, in this situation, from the recruitment of RPA to the single-stranded DNA overhangs generated by end resection. Furthermore, depletion of SMARCAD1 sensitises human cells in culture to the genotoxin camptothecin (Costelloe et al., 2012). Based on these observations, SMARCAD1, like Fun30, is regarded as a positive regulator of end resection and thereby, of DSB repair by homologous recombination (Chen et al., 2012, Costelloe et al., 2012).

1.6 KAP1

1.6.1 Domain Architecture of KAP1

The SMARCAD1 interacting protein, KAP1, is a tripartite motif-containing (TRIM) protein, as reflected in its alternative name of TRIM28 (Friedman et al., 1996, Le

Douarin et al., 1996). (Yet another alias of KAP1 is transcription intermediary factor 1 beta (TIF1β).) TRIM proteins are distinguished by their possession of an RBCC domain, which is a composite of RING, B-box and coiled-coil (RBCC) domains (Hatakeyama, 2011). Additionally, KAP1 has a centrally located HP1-box, through which it interacts with HP1, and a C-terminal PHD-bromodomain (Figure 1.15) (Friedman et al., 1996, Le Douarin et al., 1996). The closely related proteins, TIF1α (TRIM24) and TIF1γ (TRIM33), share the domain architecture of KAP1. Indeed, the RBCC domains and PHD-bromodomains of KAP1 and TIF1α share 40% and 34% sequence identity respectively; likewise, a similar level of homology occurs between KAP1 and TIF1γ (Iyengar and Farnham, 2011).

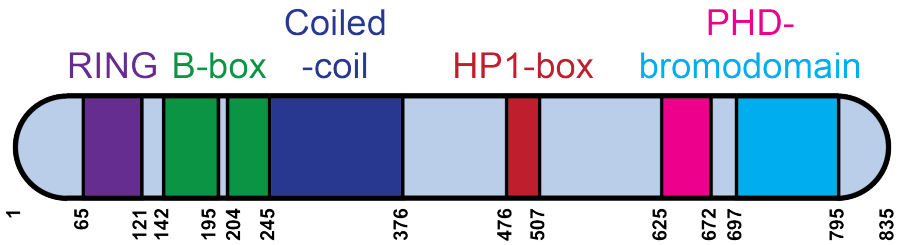


Figure 1.15 – Domain Architecture of Human KAP1
This schematic illustrates the main features of the human KAP1 protein.

RING finger domains are zinc-finger domains that frequently possess E3 ubiquitin ligase activity, though a substantial minority are catalytically inactive (Deshaies and Joazeiro, 2009). No physiological ubiquitylation targets of KAP1's RING finger domain have yet been identified; therefore, it remains unknown whether this KAP1 domain functions as a *bona fide* E3 ubiquitin ligase. (Curiously, however, the KAP1 PHD domain functions as an E3 small ubiquitin-like modifier (SUMO) ligase that promotes the sumoylation of the KAP1 bromodomain; this aspect of KAP1 biology will be elaborated upon later in this section (Ivanov et al., 2007).)

Within the KAP1 RBCC domain are a pair of tandem B-boxes. The B-box domain is structurally related to the RING domain – a globular domain stabilized by two coordinating zinc ions (Massiah et al., 2006, Tao et al., 2008). Due to the structural similarity between B-boxes and RING domains, it has been suggested that B-boxes either possess of E3 ligase activity in their own right, or are capable of enhancing that of RING domains (Massiah et al., 2006). Neither function, however, can be definitively ascribed to B-boxes.

Figure 1.16 – Structure of the TRIM25 Coiled-Coil Domain
A. The TRIM25 coiled-coil dimer is shown in orthogonal views. For clarity, one subunit is depicted in rainbow representation, beginning with violet at the N-terminus coloured and ending with red at the C-terminus. The other subunit is shown in white. The image was taken from (Sanchez et al., 2014).
B. The coiled-coil domain may the main determinant of overall structural organization for TRIM-family proteins in general. The coiled-coil dimer acts as an extended interaction interface, with the RING and B-box domains at either end. The C-terminal part of the protein clusters in the middle, and on one side of the coiled-coil domain. This image is a recreation of one in (Sanchez et al., 2014).

The coiled-coil domains of KAP1 mediate homo-trimerization of KAP1. Moreover, recent structural characterization of the TRIM25 coiled-coil homodimer has revealed how the coiled-coil domain can serve as the cornerstone that dictates the positions of the other domains of a TRIM-family protein (Figure 1.16). The construct of TRIM25 used for crystallization encompassed the coiled-coil domain and the first half of the ensuing linker region. Each TRIM25 molecule adopted a hairpin configuration; a short arm consisting of the linker region folded back onto a long arm, formed by the coiled-coil domain proper. The TRIM25 dimerized in an antiparallel orientation, with the coiled-coil domain functioning as an extended interaction interface. Intriguingly, the coiled-coil domain is comprised of heptad (7) and hendecad (11) amino acid repeats arranged in a palindromic 7-7-7-7-11-11-11-11-7-7-7-7 pattern (where ‘7’ refers to a heptad repeat and ‘11’ a hendecad repeat). The heptad repeats at the ends form canonical left-handed supercoils that flank a central under-wound, right-handed coil arising from the hendecad repeats. Thus, the linker helices of the short arm can infiltrate the central portion of the coiled-coil domain, exploiting the under-winding to form an interdigitated four-helix bundle. The coiled-coil domain of TRIM25 thereby imposes the following positional constraints on the remaining domains – the C-terminal domains cluster in the middle, and on one side, of the coiled-coil domain, while the N-terminal RING and B-box domains cap the ends. Multiple sequence alignment confirms that the coiled-coil domain of KAP1 is composed of a similar combination of heptad and hendecad repeats (Sanchez et al., 2014). Hence, it is plausible that KAP1 is organized structurally in a similar manner to TRIM25; however, in the absence of a substantial corpus of solved structures of TRIM proteins, it is unknown whether the structure of TRIM25 can be regarded as being representative of the TRIM family in general.

Located within the central region of the KAP1 protein, the HP1-box contains a hydrophobic pentapeptide PxVxL motif that mediates an interaction between KAP1 and the chromo-shadow domain of all three isoforms of HP1 (Nielsen et al., 1999, Ryan et al., 1999, Lechner et al., 2000). Indeed, KAP1 co-localizes with HP1, particularly around nucleoli and in pericentric heterochromatin (Ryan et al., 1999). This interaction with HP1 is essential for the ability of KAP1 to repress transcription to be fully realized (Nielsen et al., 1999). Notably, enforced recruitment KAP1 to euchromatic gene loci is sufficient to mobilize HP1 to these sites, culminates in transcriptional silencing and chromatin compaction (Ayyanathan et al., 2003). The KAP1 HP1-box interacts with HP1 with a stoichiometry of one to two; furthermore, since KAP1 exists as a homotrimer, it may be associated with a cluster of six molecules of HP1. KAP1 may, therefore, nucleate multimerization of HP1 and promote the conversion of the local chromatin structure into a heterochromatic state (Lechner et al., 2000). Given its interaction with KAP1, it is particularly intriguing to speculate that SMARCAD1 may contribute to this process. Yet, as KAP1 is present in the cell at considerable excess compared to SMARCAD1, the SMARCAD1 may function entirely independently of KAP1-mediated transcriptional repression that involves HP1.

PHD domains typically mediate interactions with histone tails, for example, enabling the specific recognition of the unmethylated H3K4me0 or tri-methylated H3K4me3 states (Patel and Wang, 2013). However, the PHD domain of KAP1 possesses neither a conserved aspartic acid residue to form an ion pair with unmethylated lysine 4, nor an aromatic cage to stabilize an interaction with tri-methylated lysine 4 through cation-π interactions (Chakravarty et al., 2009). Likewise, though most bromodomains enable interactions with acetylated histone tails, this is not true for the KAP1 bromodomain (Zeng et al., 2008, Filippakopoulos et al., 2012). One characteristic of canonical bromodomains is a hydrophobic pocket that can be occupied by acetylated histone tails; the KAP1 bromodomain, however, has a distorted pocket, thereby explaining its inability to undertake histone acetyl-lysine recognition (Zeng et al., 2008).

Instead of histone mark recognition, the KAP1 PHD-bromodomain has evolved to interact with the histone methyl-transferase SETDB1, which deposits the repressive

H3K9me3 modification, and CHD3, a catalytic subunit of the NuRD complex (Schultz et al., 2001, Schultz et al., 2002, Ivanov et al., 2007). These interactions are dependent on sumoylation of the KAP1 PHD-bromodomain as they involve SUMO-interacting motifs (SIMs) on SETDB1 and CHD3. Uniquely, the KAP1 PHD domain functions as an E3 SUMO ligase that promotes the sumoylation of KAP1's bromodomain (Ivanov et al., 2007). The SUMO moiety is highly related to ubiquitin; indeed, the process of sumoylation is analogous to that of ubiquitylation (Flotho and Melchior, 2013). The PHD domain of KAP1 binds to Ubc9, the E2 conjugating enzyme, and presumably as a consequence, the E2~SUMO thioester is oriented in the optimal position to transfer the SUMO moiety to lysine residues of the bromodomain (Ivanov et al., 2007). The PHD and bromodomain of bromodomain form one structural domain that behaves as a single functional unit – compromising the PHD-bromodomain interface reduces the ability of KAP1 to repress transcription (Zeng et al., 2008). Similarly, the ability of KAP1 to function as a transcriptional co-repressor for KRAB ZNF transcription factors is also compromised under the following scenarios: by mutating the sumoylation sites on KAP1, by jeopardizing the integrity of the PHD domain, and by inhibiting the sumoylation machinery (Ivanov et al., 2007). Collectively, these data emphasize that recruitment of SETDB1 and CHD3 by the sumoylated PHD-bromodomain is an integral facet of KAP1-mediated transcriptional repression (Schultz et al., 2001, Schultz et al., 2002, Sripathy et al., 2006, Ivanov et al., 2007, Zeng et al., 2008).

Despite its role in transcriptional repression and its association with chromatin, KAP1 does not itself contain a DNA-binding domain and is presumably recruited to chromatin by its interaction partners (Iyengar and Farnham, 2011). Nevertheless, approximately 7000 KAP1 binding sites were identified by chromatin immunoprecipitation-microarray chip (ChIP-chip) analysis (O'Geen et al., 2007). KAP1 genomic targets can be divided into two main categories. The first category consists of the 3' ends of ZNF genes, to which KAP1 is recruited via KRAB ZNF transcription factors in a RBCC domain-dependent manner. Although the recruitment of KAP1 to ZNF genes by KRAB ZNF proteins would suggest an auto-regulatory feedback loop, depletion of KAP1 does not affect the expression of the majority of these genes (Iyengar et al., 2011). Yet, a role in transcriptional regulation for KAP1 at these genes cannot be formally excluded as stable

inheritance of KAP1-induced heterochromatin has been observed, suggesting that it may be possible for direct KAP1-mediated transcriptional repression to be perpetuated in the transient absence of KAP1 (Ayyanathan et al., 2003). The second category represents targets that KAP1 binds in a KRAB ZNF protein-independent manner; it includes the promoter regions of ZNF genes and non-ZNF genes. Therefore, factors other than ZNF transcription factors contribute to the recruitment of KAP1 to chromatin; for instance, the KAP1-HP1 interaction is necessary for the efficient tethering of KAP1 to its genomic targets (Iyengar et al., 2011). Thus, to fully comprehend the biological functions of KAP1, exhaustively defining the co-factors that target KAP1 to DNA will inevitably be required.

In the context of the SMARCAD1-KAP1 complex, the factors influencing its genomic localization remain undefined. Thus, it is unknown whether SMARCAD1 can exploit the modular architecture of KAP1 to form ternary complexes specific for certain regions of the genome. For example, is it possible for the SMARCAD1-KAP1 complex to be recruited to constitutive heterochromatin by forming a ternary complex with HP1?

1.6.2 Functions of KAP1

KAP1 is essential for physiological development (Iyengar and Farnham, 2011). Indeed, KAP1 knockout mice are embryonically lethal due to a failure to undergo gastrulation, the process through which the blastula is reorganized into a tri-laminar structure primed for subsequent organogenesis (Cammas et al., 2000). Correspondingly, KAP1 is implicated in the maintenance of pluripotency in mouse ESCs, and is essential for terminal differentiation of embryonal carcinoma cells along the endodermal lineage (Cammas et al., 2004, Hu et al., 2009).

At a mechanistic level, KAP1 is best characterized as an obligatory co-repressor for KRAB ZNF transcription factors (Figure 1.17). Via its RBCC domain, KAP1 interacts as a homotrimer with the KRAB repression domain of these transcription factors (Friedman et al., 1996, Moosmann et al., 1996, Peng et al., 2000). Targeting KAP1 to euchromatic gene loci results in their transcriptional silencing

and conversion into heterochromatin (Ayyanathan et al., 2003). KAP1-mediated transcriptional repression is dependent on the recruitment of additional chromatin-associated factors including HP1, via the KAP1 HP1-box, and the H3K9 histone methyl-transferase SETDB1, which contains a SIM and interacts with the sumoylated PHD-bromodomain of KAP1 (Schultz et al., 2002, Sripathy et al., 2006, Ivanov et al., 2007). These processes are highly cooperative; for instance, sumoylation of the KAP1 bromodomain by the adjacent PHD domain is compromised in a KAP1 mutant that is unable to interact with HP1. Conversely, the sumoylated PHD-bromodomain stimulates the histone methyl-transferase activity of SETDB1 *in vitro* (Ivanov et al., 2007). The sumoylated PHD-bromodomain of KAP1 also recruits the chromatin remodeller, CHD3, which promotes local chromatin compaction (Goodarzi et al., 2011). Furthermore, since HP1 recognizes the H3K9me3 mark, stable genomic localization of KAP1 and spreading of the H3K9me3 mark may be mediated by the HP1-KAP1 interaction (Jacobs and Khorasanizadeh, 2002). In summary, by exploiting its modular architecture, KAP1 simultaneously recruits several effectors of chromatin structure to the same region of the genome. The amalgamation of the activities of these proteins is the conversion of the local chromatin structure into a form recapitulating the principal features of heterochromatin – highly compacted chromatin associated with HP1 and H3K9me3 – thereby ensuring robust transcriptional repression (Ayyanathan et al., 2003, Sripathy et al., 2006, Ivanov et al., 2007, Goodarzi et al., 2011).

Figure 1.17 – Model of KAP1-Mediated Transcriptional Repression
Due to its modular architecture, KAP1 is able to use its HP1-box to recruit HP1 to chromatin, while simultaneously using its sumoylated bromodomain to interact with the NuRD complex and the H3K9 methyl-transferase, SETDB1. These interactions enable KAP1 to convert euchromatic gene loci into transcriptionally silent, compacted heterochromatin containing HP1 and bearing the repressive H3K9me3 modification. The figure was adapted from (Ivanov et al., 2007).

Aside from transcriptional repression, KAP1 is essential for ataxia telangiectasia mutated (ATM) kinase-dependent repair of heterochromatic DSBs (Goodarzi et al., 2008). In response to DNA damage, ATM phosphorylates KAP1 at serine 824 (Ziv et al., 2006). Phosphorylation of this residue disrupts the interaction between the sumoylated KAP1 bromodomain and a SIM on CHD3. Dissociation of the CHD3

chromatin remodeller results in relaxation of the local chromatin environment, which is reflected in an increase in micrococcal nuclease accessibility (Goodarzi et al., 2011). The reduction in chromatin compaction is presumably required to render the region permissive to the recruitment of the DNA repair machinery (Ziv et al., 2006, Goodarzi et al., 2008, Goodarzi et al., 2011).

It is curious that KAP1 interacts with multiple chromatin remodellers – CHD3 and SMARCAD1. However, it is unlikely that they are functionally redundant since they bear little similarity to one another in terms of domain architecture and complex composition. Indeed, it is entirely possible that CHD3 and SMARCAD1 function independently of each other – KAP1 is present in considerable excess in the cell relative to SMARCAD1, which consequently, would only be able to interact with a relatively small proportion of KAP1 molecules. Functional interactions between the two chromatin remodellers may, therefore, be limited in scope. On this basis, it is also possible that the canonical role for KAP1 in KRAB ZNF protein-mediated transcriptional repression may be irrelevant for its function with SMARCAD1. Virtually nothing is known about the SMARCAD1-KAP1 interaction, much less about the functional interplay between the two proteins. Thus, we felt that it was essential to investigate the effect of KAP1 on the catalytic activity of SMARCAD1 *in vitro*, and our findings are discussed in Chapter 6.

1.7 CUE Domains

The coupling of ubiquitin conjugation to endoplasmic reticulum degradation (CUE) domain earned its unusual epithet from the protein in which it as initially identified, the *S. cerevisiae* Cue1p protein, which recruits the ubiquitin E2 ligase, Ubc7p, to the endoplasmic reticulum membrane to promote proteasomal degradation (Biederer et al., 1997). A small protein motif of approximately 40 amino acids, the CUE domain is moderately well conserved across eukaryotes ranging from fungi to metazoans (Ponting, 2000). CUE domains adopt a globular configuration comprised of a bundle of three anti-parallel α-helices, and generally act as ubiquitin-binding domains (UBDs) (Shih et al., 2003, Hicke et al., 2005).

1.7.1 CUE Domains are Ubiquitin-Binding Domains

Ubiquitin is a small, stable protein of 76 residues that displays high evolutionary conservation – aside for three conservative amino acid substitutions, its sequence is invariant between yeast and man (Komander and Rape, 2012). Ubiquitin assumes a compact β-grasp fold, consisting of a mixed five-strand β-sheet covered by an α-helix, with an exposed C-terminus (Vijay-Kumar et al., 1987). Ubiquitylation is a post-translational modification that involves the covalent linkage of the C-terminus of ubiquitin to the ε-amino group of a lysine residue via an isopeptide linkage. Uniquely, conjugated ubiquitin moieties can themselves be modified, for example, by further ubiquitylation at any of its six lysine residues. While numerous arrangements of ubiquitin chains have been described, the two canonical forms involve either lysine 48 or lysine 63 linkages, with the former displaying a compact conformation and the latter an open one (Komander and Rape, 2012). Although ubiquitylation was initially regarded as the PTM that targets proteins for proteasomal degradation, it is now undisputable that ubiquitylation is also intimately implicated as a signalling device in a diverse range of biological processes, ranging from transcriptional regulation to immunity (Husnjak and Dikic, 2012).

For ubiquitylation to regulate cellular processes, it is necessary that the ubiquitin moiety can be recognized. Indeed, cells have evolved a diverse set of UBDs, most of which bind to a large hydrophobic patch on the surface of the ubiquitin moiety, centred on the leucine 8, isoleucine 44 and valine 70 residues (Husnjak and Dikic, 2012). This is true of CUE domains, which contact ubiquitin using an interface formed from two of its helices (α1 and α3). Specifically, the interaction interface of the CUE domain features a hydrophobic patch complementary to the L8-I44-V70 hydrophobic patch of ubiquitin (Figure 1.18). The ubiquitin-CUE domain interaction is further stabilized by electrostatic contacts along the rim of the interaction interface (Kang et al., 2003, Prag et al., 2003, Liu et al., 2012). Multiple-sequence alignment of CUE domains defines two highly conserved elements – a methionine-phenylalanine-proline (MFP) motif in the loop between α1 and α2 helices, and a di-leucine repeat at the C-terminal end of the α3 helix (Ponting, 2000, Prag et al., 2003). Many of these conserved residues contribute significantly to establishing the hydrophobic character of the CUE domain interaction interface, explaining the

correlation between their presence and high affinity ubiquitin binding (Kang et al., 2003, Prag et al., 2003, Shih et al., 2003, Liu et al., 2012). The phenylalanine residue of the MFP motif also stabilizes the hydrophobic core of the CUE domain; therefore, substitution of phenylalanine with an alanine residue in the first CUE domain of the CUE2 protein or in the gp78 CUE domain disrupts protein folding (Kang et al., 2003, Liu et al., 2012).

The interaction interface between the CUE domain and ubiquitin is limited, only spanning an area of approximately 400Å^2. Consequently, even with fully intact MFP and LL motifs, CUE domains typically recognize ubiquitin with only modest affinity (Kang et al., 2003, Prag et al., 2003, Liu et al., 2012). Yet, CUE domains vary considerably in their affinity for ubiquitin, with dissociation constants ranging from approximately 20 to 160μM (Shih et al., 2003). Some of this variation could be explained by the ability of some CUE domains to function as a dimer to bind ubiquitin with higher affinity. For instance, as a dimer, the CUE domain of the yeast Vps9p protein is able to form additional contacts between ubiquitin and the typically unutilized α2 helix of the CUE domain (Prag et al., 2003).

As indicated by the dissociation constants listed above, the ubiquitin-UBD interaction is typically weak. The biological significance of this phenomenon is unclear; however, it is speculated that the relatively low affinity is enables binding to be rapidly reversed, as befitting the role of ubiquitin as a broadly applicable regulatory mechanism. It is also plausible that ubiquitin-UBD interactions are weak to accommodate high cellular concentrations of free ubiquitin, which in mammalian cells, is estimated to be approximately 10μM. If free ubiquitin were to interact with UBDs with high affinity, UBDs may be irreversibly occupied by ubiquitin, precluding their interaction with ubiquitylated partner proteins (Hicke et al., 2005). Importantly, ubiquitin-binding domains also serve merely as secondary interaction domains ensuring specificity, strengthening an otherwise weak interaction between two proteins, so that the ubiquitylated version can be strongly selected for interaction.

Figure 1.18 – Structure of the First CUE Domain of Cue2p Bound to Ubiquitin
A. In this structure, ubiquitin is coloured in green and the first CUE domain of the Cue2p protein in violet. Two helices (α1 and α3) form a hydrophobic surface that packs against

the hydrophobic patch centred upon the L8-I44-V70 residues of ubiquitin. This structure was retrieved from the PDB as 1OTR.

B. The interaction interface between the CUE domain and ubiquitin is shown here. The side chains of residues that contribute to the interaction are depicted here in stick representation, and important ubiquitin residues are labelled. This image was taken from (Kang et al., 2003).

Aside for ubiquitin, the only other known ligand for CUE domains is the ubiquitin-homology (UbH) domain. UbH domains are integral protein domains that adopt the β-grasp ubiquitin superfold conformation. As UbH domains bear structural resemblance to ubiquitin, they are frequently capable of interacting with UBDs, including CUE domains (Grabbe and Dikic, 2009). For example, Def1, a budding yeast protein that functions as a bridging factor, exploits a UbH-CUE domain interaction to recruit the Elongin-Cullin (Ela1-Elc1) E3 ubiquitin ligase complex to RNAPII via the Def1 CUE domain and an UbH domain in Ela1. While the UbH-CUE domain interaction probably involves the hydrophobic patches of both domains, specificity is probably conferred by additional contacts as only the Def1 CUE domain, but not several other UBDs tested, was able to enrich for the Elongin-Cullin complex from yeast extracts (Wilson et al., 2013).

1.8 Aims of this Thesis

Although important roles in development have been described for CHD1 and CHD5, a cogent mechanistic explanation for these functions has yet to be offered. We hypothesized that like the BAF complex, CHD1 and CHD5 associate with hitherto unidentified tissue-specific complex subunits, which confer upon them the ability to regulate developmental processes. Thus, I sought to purify endogenous CHD1 from mouse ESCs to identify complex subunits or interactors that are specific to the pluripotent cell state by mass spectrometry. Likewise, I aimed to purify the native CHD5 chromatin remodelling complex from mouse brains such as to determine its complex composition. Though no novel tissue-specific complex subunits were identified from detailed mass spectrometry analysis, the purification of CHD1 and CHD5 are described in Chapters 3 and 4 respectively.

Due to the largely negative results associated with the characterization of these scarce CHD proteins, the remainder of this thesis focuses upon our biochemical

characterization of the poorly studied chromatin remodeller, SMARCAD1, which is implicated in the maintenance of heterochromatin and double-stranded DNA break repair. To better understand the regulation of this enzyme, we focussed upon the interplay between SMARCAD1 and KAP1, between which a constitutive stoichiometric protein complex is formed. First, we wanted to determine whether KAP1 affects the catalytic activity of SMARCAD1 *in vitro*. Following reconstitution of the SMARCAD1-KAP1 complex with purified recombinant proteins, we were able to demonstrate that KAP1 acts as a negative regulator of the nucleosome-stimulated ATPase activity of SMARCAD1, as discussed in Chapter 6. Next, we sought to characterize the SMARCAD1-KAP1 interaction by identifying the regions of each protein involved, and by determining the nature of the interaction. Thus, Chapter 7 pertains to our findings that the interaction is a direct protein-protein interaction that involves the first SMARCAD1 CUE domain and the RBCC domain of KAP1, but does not require ubiquitylation of KAP1. As a ubiquitin-homology domain cannot be identified within KAP1, an attempt to determine the co-crystal structure of the SMARCAD1-KAP1 interaction interface is also described.

Chapter 2. Materials & Methods

2.1 Buffers, Media and Solutions

2.1.1 Bacterial Media

2.1.1.1 Lysogeny Broth (LB), pH 7.0

1% (w/v) bacto-tryptone
0.5% (w/v) yeast extract
1% (w/v) NaCl
± 100µg/mL ampicillin (in 50% (v/v) ethanol)
± 50µg/mL kanamycin
± 50µg/mL spectinomycin
± 100µg/mL streptomycin
± 34µg/mL chloramphenicol (in ethanol)
LB was prepared by the Media & Cell Services core facility (LRI).

2.1.1.2 NZY Broth, pH 7.0

10mg/mL yeast extract
5mg/mL NaCl
2mg/mL glucose
16mg/mL NZ-Amine A

2.1.1.3 SOC Media, pH 7.0

2% (w/v) bacto-tryptone
0.5% (w/v) yeast extract
10mM NaCl
2.5mM KCl
10mM $MgCl_2$
10mM $MgSO_4$
20mM glucose

2.1.2 Mammalian Tissue Culture Media

2.1.2.1 Dulbecco's Modified Eagle's Medium (DMEM)

DMEM was obtained from Gibco® (Thermo Fisher Scientific) and supplemented with 10% (v/v) foetal bovine serum (labtech.com) and penicillin (20U/mL)/streptomycin (100µg/mL), supplied as a 100X stock solution (Sigma).

2.1.2.2 ESC Medium

DMEM supplemented with the following components:
2.4mM L-glutamine
15% (v/v) foetal bovine serum
50µM β-mercaptoethanol (β-ME)
10ng/mL leukaemia inhibitory factor (LIF)

2.1.2.3 Freezing Medium

DMEM supplemented with the following components:
20% (v/v) foetal bovine serum
10% (v/v) DMSO

2.1.2.4 Opti-MEM Reduced Serum Medium

Obtained from Gibco® (Thermo Fisher Scientific)

2.1.2.5 Trypsin/EDTA Solution, pH 7.0-7.6

0.5g/L trypsin
0.2g/L EDTA
Phenol red
Obtained from Sigma-Aldrich

2.1.3 General Solutions

2.1.3.1 1X SDS-PAGE Loading Buffer

7.5% (v/v) glycerol
70mM Tris pH 6.8
0.25% (w/v) SDS
0.125M Dithiothreitol (DTT)
Bromophenol blue

2.1.3.2 4-6X SDS-PAGE Loading Buffer

30% (v/v) glycerol
0.28M Tris pH 6.8
1% (w/v) SDS
0.5M DTT
Bromophenol blue

2.1.3.3 6X DNA Loading Buffer

60% (v/v) glycerol
10mM Tris pH 7.5
60mM EDTA
0.25% (w/v) Orange G

2.1.3.4 100X Protease Inhibitor Cocktail (in Ethanol)

28.4µg/ml leupeptin
137µg/ml pepstatin A
17mg/ml phenylmethylsulfonyl fluoride (PMSF)
33mg/ml benzamindine

2.1.3.5 Lysozyme Reconstitution Buffer

10mM Tris, pH 7.50
10% (v/v) glycerol

2.1.3.6 Phosphate Buffered Saline (PBS), pH 7.50

137mM NaCl
2.7mM KCl
10mM Na_2HPO_4
2mM NaH_2PO_4
PBS was prepared by the Media & Cell Services core facility (LRI).

2.1.3.7 PBS-Tween

137mM NaCl
2.7mM KCl
10mM Na_2HPO_4
2mM NaH_2PO_4
0.05% (v/v) Tween 20

2.1.3.8 SDS-PAGE Running Buffer

SDS-PAGE running buffer was prepared by dilution of 20X XT MOPS or MES Running Buffer stock solutions (Bio-Rad).

2.1.3.9 TBS-Tween

20mM Tris pH7.50 at room temperature (RT)
100mM NaCl
0.1% (v/v) Tween 20

2.1.3.10 Tris-Borate-EDTA (TBE) Buffer

89mM Tris-Cl
89mM boric Acid
2mM EDTA
1X TBE was prepared by dilution of a 10X stock buffer.

2.1.3.11 Trypsin Reconstitution Buffer

50mM acetic acid
10% (v/v) glycerol

2.1.3.12 Transfer Buffer

25mM Tris-base
192mM glycine
20% (v/v) methanol
0.02% (v/v) SDS

2.1.3.13 Triton Lysis Buffer, pH 7.50

50mM Tris
150mM NaCl
1mM EDTA
1% (v/v) Triton X-100

2.1.4 Mammalian Purification Buffers

2.1.4.1 WCE-150 Buffer, pH 7.90 at 4°

20mM Hepes
150mM NaCl
1mM EDTA
0.5% (v/v) NP-40

2.1.4.2 B-100 Buffer, pH 7.90 at 4°C

10mM Hepes
10mM sodium phosphate
100mM NaCl

2.1.4.3 R-150N Buffer, pH 7.80 at 4°C

50mM Tris
150mM NaCl
1mM EDTA
10μM $ZnCl_2$
RG-150N buffer also has 40% (v/v) propylene glycol

2.1.4.4 R-500Am Buffer, pH 7.80 at 4°C

50mM Tris
500mM $(NH)_2SO_4$
1mM EDTA
10μM $ZnCl_2$
RG-150N buffer also has 40% (v/v) propylene glycol

2.1.4.5 CHD-150 Buffer, pH 7.50 at 4°C

50mM Tris
150mM NaCl
10% (v/v) glycerol
10mM EDTA
1mM DTT

2.1.4.6 CHD-750 Buffer, pH 7.50 at 4°C

50mM Tris

750mM NaCl
10% (v/v) glycerol
10mM EDTA
1mM DTT

2.1.4.7 CHD-50 Buffer, pH 7.50 at 4°C

50mM Tris
50mM NaCl
10% (v/v) glycerol
10mM EDTA
1mM DTT

2.1.4.8 TEV Buffer, pH 8.00 at 4°C

50mM Tris
100mM NaCl
10% (v/v) glycerol
1mM sodium citrate
1mM DTT
0.01% (v/v) NP-40

2.1.4.9 CL-50 Buffer, pH7.9.0 at 4°C

20mM Hepes
50mM NaCl
10% (v/v) glycerol
2mM $MgCl_2$
1mM DTT
0.01% (v/v) NP-40

2.1.4.10 C-100 Buffer, pH7.9.0 at 4°C

50mM Hepes
100mM NaCl
10% (v/v) glycerol
1mM DTT
0.01% (v/v) NP-40
C-100Ca also contains 2mM $CaCl_2$

2.1.4.11 C-750 Buffer, pH7.9.0 at 4°C

50mM Hepes
750mM NaCl
10% (v/v) glycerol
10mM EDTA
1mM DTT
0.01% (v/v) NP-40

2.1.4.12 C-50 Buffer, pH7.9.0 at 4°C

50mM Hepes
50mM NaCl
10% (v/v) glycerol
10mM EDTA
1mM DTT
0.01% (v/v) NP-40

2.1.4.13 C-2000 Buffer, pH7.9.0 at 4°C

50mM Hepes
2M NaCl
10% (v/v) glycerol
10mM EDTA

1mM DTT
0.01% (v/v) NP-40

2.1.5 Bacterial Purification Buffers

2.1.5.1 HL-500 Buffer, pH 7.90 at 4°C

20mM Tris
500mM NaCl
10mM imidazole
10% (v/v) glycerol
0.1% (v/v) NP-40
5mM β-ME
HL-500Zn also contains 50μM $ZnSO_4$

2.1.5.2 HW-750 Buffer, pH 7.90 at 4°C

20mM Tris
750mM NaCl
10mM imidazole
10% (v/v) glycerol
5mM β-ME
HW-750Zn also contains 50μM $ZnSO_4$

2.1.5.3 HW-100 Buffer, pH 7.90 at 4°C

20mM Tris
100mM NaCl
10mM imidazole
10% (v/v) glycerol
5mM β-ME
HW-100Zn also contains 50μM $ZnSO_4$

2.1.5.4 HE-300 Buffer, pH 7.90 at 4°C

20mM Tris
100mM NaCl
300mM imidazole
10% (v/v) glycerol
5mM β-ME
HE-300Zn also contains 50μM $ZnSO_4$

2.1.5.5 P-50 Buffer, pH 7.50 at 4°C

10mM sodium phosphate
50mM NaCl
10% (v/v) glycerol
5mM β-ME

2.1.5.6 P-100 Buffer, pH 7.50 at 4°C

10mM sodium phosphate
100mM NaCl
10% (v/v) glycerol
5mM β-ME
P-100Zn also contains 50μM $ZnSO_4$

2.1.5.7 P-1000 Buffer, pH 7.50 at 4°C

10mM sodium phosphate
1M NaCl
10% (v/v) glycerol
5mM β-ME
P-1000Zn also contains 50μM $ZnSO_4$

2.1.5.8 Q-100 Buffer, pH 7.90 at 4°C

10mM Tris
100mM NaCl
10% (v/v) glycerol
5mM β-ME
Q-100Zn also contains 50μM $ZnSO_4$

2.1.5.9 Q-1000 Buffer, pH 7.90 at 4°C

10mM Tris
1M NaCl
10% (v/v) glycerol
5mM β-ME
Q-1000Zn also contains 50μM $ZnSO_4$

2.1.5.10 HL-200Zn Buffer, pH 7.90 at 4°C

20mM Tris
200mM NaCl
10mM imidazole
10% (v/v) glycerol
0.1% (v/v) NP-40
5mM β-ME
50μM $ZnSO_4$

2.1.5.11 HW-200Zn Buffer, pH 7.90 at 4°C

20mM Tris
200mM NaCl
10mM imidazole
10% (v/v) glycerol
5mM β-ME

50μM $ZnSO_4$

2.1.5.12 GF-150Zn Buffer, pH 7.90 at 4°C

10mM Tris
150mM NaCl
50μM $ZnSO_4$
2mM DTT

2.1.5.13 Ulp1 Lysis Buffer, pH 7.90 at 4°C

20mM Tris
500mM NaCl
10mM imidazole
5mM β-ME

2.1.5.14 Ulp1 Wash Buffer 1, pH 7.90 at 4°C

20mM Tris
750mM NaCl
20mM imidazole
5mM β-ME

2.1.5.15 Ulp1 Wash Buffer 2, pH 7.90 at 4°C

20mM Tris
500mM NaCl
20mM imidazole
5mM β-ME

2.1.5.16 Ulp1 Elution Buffer, pH 7.90 at 4°C

20mM Tris

500mM NaCl
200mM imidazole
5mM β-ME

2.1.5.17 Ulp1 GF Buffer, pH 7.90 at 4°C

25mM Tris
500mM NaCl

2.1.5.18 GST-L Buffer, pH 7.90 at 4°C

20mM Tris
100mM NaCl
1mM EDTA
10% (v/v) glycerol
0.1% (v/v) NP-40

2.1.5.19 GST-W500 Buffer, pH 7.90 at 4°C

10mM Tris
500mM NaCl
10% (v/v) glycerol
5mM β-ME

2.1.5.20 GST-W50 Buffer, pH 7.90 at 4°C

10mM Tris
50mM NaCl
10% (v/v) glycerol
5mM β-ME

2.1.5.21 GST-100 Buffer, pH 7.90 at 4°C

10mM Tris
100mM NaCl
10% (v/v) glycerol
5mM β-ME

2.1.5.22 STE Buffer

10mM Tris, pH 8.0
1mM EDTA
100mM NaCl

2.1.5.23 MD Wash Buffer 1

1X PBS
450mM NaCl
10% (v/v) glycerol
0.1mM EDTA
0.1% (v/v) Triton™ X-100
2mM DTT

2.1.5.24 MD Wash Buffer 2

50mM sodium phosphate
50mM NaCl
10% (v/v) glycerol
1mM β-ME
0.2% (v/v) Triton™ X-100

2.1.6 Biochemical Assay Buffers

2.1.6.1 5X SK Reconstitution Buffer, pH7.50 at RT

50mM sodium phosphate
1M NaCl
50% (v/v) glycerol
0.5% (v/v) NP-40

2.1.6.2 P-200 GF Buffer, pH7.50 at 4°C

10mM sodium phosphate
200mM NaCl
2mM DTT

2.1.6.3 5X ATPase Buffer

50mM Tris pH7.90
0.5mM $MgCl_2$
2.5mM DTT
0.4mg/mL BSA

2.1.6.4 TLC Running Buffer

1M formic acid
0.3M LiCl

2.1.6.5 5X SK Binding Buffer

50mM Tris, pH7.50 at RT
750mM NaCl
50% (v/v) glycerol
0.05% (v/v) NP-40
250µM $ZnSO_4$

2.1.6.6 SK-200 Buffer

10mM Tris, pH7.50 at RT

200mM NaCl

10% (v/v) glycerol

0.01% (v/v) NP-40

50μM $ZnSO_4$

2.1.6.7 GST-L-Zn Buffer, pH7.90 at 4°C

20mM Tris

100mM NaCl

10% (v/v) glycerol

0.1% (v/v) NP-40

50μM $ZnSO_4$

5mM β-ME

2.1.6.8 10X Trypsin Buffer

200mM Tris, pH 7.40

500mM NaCl

10mM $CaCl_2$

20mM DTT

2.2 DNA Techniques

2.2.1 Plasmids

Plasmid	Description	Cloning Sites	Selection Markers	Source
p601	Widom 601 sequence in a pBluescript vector	Unknown	Amp	DM
pCDF-Duet1	Bacterial expression vector for coexpression of 2 ORFs	N/A	Strep/Spec	Novagen
pCDF-Duet1-SUMO	Smt3 (SUMO) cloned into pCDF-Duet1; enhances solubility of recombinant proteins; untagged	NcoI & BamHI	Strep/Spec	PL
pCDF-SUMO-KAP1 S33-K434	KAP1 S33-K434 cloned into pCDF-Duet1-SUMO	BamHI & KpnI	Strep/Spec	ML

pCDF-SUMO-KAP1 D202-K434	KAP1 D202-K434 cloned into pCDF-Duet1-SUMO	BamHI & KpnI	Strep/Spec	ML
pcDNA4/TO	Mammalian expression vector driven by the CMV promoter; for doxycycline inducible protein expression, using the T-REx system	N/A	Amp	Invitrogen
pSMARCAD1-716R/TO	SMARCAD1 cDNA with N-terminal FLAG tag cloned into pcDNA4/TO; resistant to shRNA knockdown by pGIPZ V2LHS-51716	BamHI & XhoI	Amp	HW
pSMARCAD1-716R-CUE1mt,2mt/TO	Derivative of pSMARCAD1-716R/TO; both CUE domains are point mutated (F169A, P170A, L195A, L196A, F263A, P264A, L289A, K290A)	BamHI & XhoI	Amp	HW
pcDNA5/FRT	Mammalian expression vector driven by the CMV promoter; allows establishment of stable cell lines with Flp-In system	N/A	Amp	Invitrogen
pcDNA5-FTH/FRT	Derivative of pcDNA5/FRT that adds C-terminal FLAG-TEV-HA (FTH) tag	Kpn1 & BamHI	Amp	ML
pCHD1-FTH/FRT	CHD1 cDNA cloned into pcDNA5-FTH/FRT; C-terminal FTH tag	HindIII & KpnI	Amp	ML
pCHD5-FTH/FRT	CHD5 cDNA cloned into pcDNA5-FTH/FRT; C-terminal FTH tag	HindIII & KpnI	Amp	ML
pHTF-cDNA5/FRT	Derivative of pcDNA5/FRT that adds N-terminal HA-TEV-FLAG (HTF) tag	HindIII & KpnI	Amp	ML
pHTF-CHD5/FRT	CHD5 cDNA cloned into pHTF-cDNA5/FRT; N-terminal HTF tag	KpnI & NotI	Amp	ML
pKAP1-HA/FRT	KAP1 cDNA with C-terminal HA tag cloned into pcDNA5/FRT	NheI & HindIII	Amp	ML
pKAP1-10×UbMt/FRT	Derivative of pKAP1-HA/FRT; 10 ubiquitylation sites (lysines) mutated to arginine residues	PasI & KasI	Amp	ML
pF-SMARCAD1/FRT	SMARCAD1 cDNA with N-terminal FLAG tag cloned into pcDNA5/FRT	BamHI & XhoI	Amp	HW
pF-SMARCAD1-CUE1mt,2/FRT	Derivative of pSMARCAD1-F/FTH; 1st CUE domain is point mutated (F169A, P170A, L195A, L196A)	BamHI & XhoI	Amp	HW
pF-SMARCAD1-CUE1,2mt/FRT	Derivative of pSMARCAD1-F/FTH; 2nd CUE domain is point mutated (F263A, P264A, L289A, K290A)	BamHI & XhoI	Amp	HW
pF-SMARCAD1-CUE1mt,2mt/FRT	Derivative of pSMARCAD1-F/FTH; both CUE domains are point mutated (F169A, P170A, L195A, L196A, F263A, P264A, L289A, K290A)	BamHI & XhoI	Amp	HW
pIRES-puro	Mammalian expression vector driven by the CMV promoter	N/A	Amp	Clontech
pMCS-IRES-puro	Derivative of pIRES-puro with expanded multiple cloning site	EcoRV & NotI	Amp	ML
pCHD5F-IRES-puro	CHD5 cDNA with C-terminal FLAG tag cloned into pIRES-puro	NotI & EcoRI	Amp	ML
pKIAA1045HA-IRES-puro	KIAA1045 cDNA with C-terminal HA tag cloned into pIRES-puro	NotI & EcoRI	Amp	ML
pET21b	Bacterial expression vector; C-terminal 6×His tag; NdeI (CAT**ATG**) in MCS	N/A	Amp	Novagen
pET21b-KAP1 S33-K434	Untagged KAP1 S33-K434 cloned into pET21b	NdeI & EcoRI	Amp	ML
pET21b-KAP1 D202-K434	Untagged KAP1 D202-K434 cloned into pET21b	NdeI & EcoRI	Amp	ML
pET28a	Bacterial expression vector; N-terminal 6×His tag	N/A	Kan	Novagen
pET28a-SUMO	Smt3 (SUMO) cloned into pET28a; enhances solubility of recombinant proteins; N-terminal 6×His tag	NheI & BamHI	Kan	PC
pET28a-SUMO-F-Fun30	Fun30 cDNA with N-terminal FLAG tag cloned into pET28a-SUMO	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-Fun30-CUEmt	Derivative of pET28a-SUMO-F-Fun30; CUE domain is point mutated (F82A, P83A)	BamHI & EcoRI	Kan	ML

pET28a-SUMO-KAP1	KAP1 cDNA cloned into pET28a-SUMO	BamHI & EcoRI	Kan	ML
pET28a-SUMO-KAP1-HA	KAP1 cDNA with C-terminal FLAG tag cloned into pET28a-SUMO	BamHI & EcoRI	Kan	ML
pET28a-SUMO-HA-KAP1	KAP1 cDNA with N-terminal FLAG tag cloned into pET28a-SUMO	BamHI & EcoRI	Kan	ML
pET28a-SUMO-HA-KAP1 RBCC	KAP1 1-388 with N-terminal HA tag cloned into pET28s-SUMO vector	BamHI & EcoRI	Kan	ML
pET28a-SUMO-HA-KAP1 ΔRBCC	KAP1 389-835 with N-terminal HA tag cloned into pET28s-SUMO vector	BamHI & EcoRI	Kan	ML
pET28a-SUMO-HA-KAP1 PB	KAP1 624-835 with N-terminal HA tag cloned into pET28s-SUMO vector	BamHI & EcoRI	Kan	ML
pET28a-SUMO-HA-KAP1 ΔPB	KAP1 1-623 with N-terminal HA tag cloned into pET28s-SUMO vector	BamHI & EcoRI	Kan	ML
pET28a-SUMO-HA-KAP1 HP1	KAP1 389-623 with N-terminal HA tag cloned into pET28s-SUMO vector	BamHI & EcoRI	Kan	ML
pET28a-SUMO-KAP1 S33-K434	KAP1 S33-K434 cloned into pET28a-SUMO vector; soluble but purified protein sticks to spin concentrator	BamHI & EcoRI	Kan	ML
pET28a-SUMO-KAP1 D202-K434	KAP1 D202-K434 cloned into pET28a-SUMO vector; insoluble	BamHI & EcoRI	Kan	ML
pET28a-SUMO-KAP1 L592-P835	KAP1 L592-P835 (PHD-bromodomain) cloned into pET28a-SUMO vector	BamHI & EcoRI	Kan	ML
pET28a-SUMO-SMARCAD1	SMARCAD1 cDNA cloned into pET28a-SUMO	BamHI & EcoRI	Kan	ML
pET28a-SUMO-SMARCAD1-F	SMARCAD1 cDNA with C-terminal FLAG tag cloned into pET28a-SUMO; FLAG tag inaccessible to M2 agarose	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-SMARCAD1	SMARCAD1 cDNA with N-terminal FLAG tag cloned into pET28a-SUMO	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-SMARCAD1 K528A	Derivative of pET28a-SUMO-F-SMARCAD1; point mutation in Walker A motif (i.e. 'ATPase dead')	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-SMARCAD1-CUE1mt,2	Derivative of pET28a-SUMO-F-SMARCAD1; 1st CUE domain is point mutated (F169A, P170A, L195A, L196A)	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-SMARCAD1-CUE1,2mt	Derivative of pET28a-SUMO-F-SMARCAD1; 2nd CUE domain is point mutated (F263A, P264A, L289A, K290A)	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-SMARCAD1-CUE1mt,2mt	Derivative of pET28a-SUMO-F-SMARCAD1; both CUE domains are point mutated (F169A, P170A, L195A, L196A, F263A, P264A, L289A, K290A)	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-CUE1,2 (S95-N347)	S95-N347 (CUE domains) of SMARCAD1 with N-terminal FLAG tag cloned into pET28a-SUMO	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-CUE1mt,2	Derivative of pET28a-SUMO-F-CUE1,2; 1st CUE domain is point mutated (F169A, P170A, L195A, L196A)	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-CUE1,2mt	Derivative of pET28a-SUMO-F-CUE1,2; 2nd CUE domain is point mutated (F263A, P264A, L289A, K290A)	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-CUE1mt,2mt	Derivative of pET28a-SUMO-F-CUE1,2; both CUE domains are point mutated (F169A, P170A, L195A, L196A, F263A, P264A, L289A, K290A)	BamHI & EcoRI	Kan	ML
pET28a-SUMO-CUE1,2	Untagged CUE domains of SMARCAD1 (S95-N347) cloned into pET28a-SUMO	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-CUE1 (S95-E237)	S95-E237 (1st CUE domain) of SMARCAD1 with N-terminal FLAG tag cloned into pET28a-SUMO	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-CUE1mt	Derivative of pET28a-SUMO-F-CUE1; 1st CUE domain is point mutatted (F169A, P170A, L195A, L196A)	BamHI & EcoRI	Kan	ML
pET28a-SUMO-CUE1 N142-R206	N142-R206 of SMARCAD1 cloned into pET28a-SUMO; trypsin-resistant fragment of	BamHI & EcoRI	Kan	ML

	1st CUE domain			
pET28a-SUMO-CUE1 N116-R206	N116-R206 (1st CUE domain) of SMARCAD1 cloned into pET28a-SUMO	BamHI & EcoRI	Kan	ML
pET28a-SUMO-CUE1 N142-E237	N142-E237 (1st CUE domain) of SMARCAD1 cloned into pET28a-SUMO	BamHI & EcoRI	Kan	ML
pET28a-SUMO-CUE1 N116-E237	N116-E237 (1st CUE domain) of SMARCAD1 cloned into pET28a-SUMO	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-CUE2 (E238-N347)	E238-N347 (2nd CUE domain) of SMARCAD1 with N-terminal FLAG tag cloned into pET28a-SUMO	BamHI & EcoRI	Kan	ML
pET28a-SUMO-F-CUE2mt	Derivative of pET28a-SUMO-F-CUE2; 2nd CUE domain is point mutated (F263A, P264A, L289A, K290A)	BamHI & EcoRI	Kan	ML
pET42a	Bacterial expression vector; N-terminal GST & 6×His tags; NdeI (CAT**ATG**) in MCS	N/A	Kan	Novagen
pET42a-Ub	Untagged Ubiquitin cDNA cloned into pET42a	NdeI & EcoRI	Kan	ML
pET42a-Ub I44A	Derivative of pET42a-Ub; I44A mutation compromises binding	NdeI & EcoRI	Kan	ML
pGEX5X-TRIM28	KAP1 cDNA cloned into bacterial expression vector; N-terminal GST tag	Unknown	Amp	(Liang et al., 2011)
pGEX5X-TRIM28 ΔBamHI	Derivative of pGEX5X-TRIM28; BamHI site in KAP1 cDNA destroyed	Unknown	Amp	ML
pGEX6P1	Bacterial expression vector; N-terminal GST tag	N/A	Amp	GE Healthcare
pGEX6P1-KAP1	KAP1 cloned into pGEX6P1	BamHI & EcoRI	Amp	ML
pGEX6P1-KAP1 S33-K434	KAP1 S33-K434 cloned into pGEX6P1	BamHI & EcoRI	Amp	ML
pGEX6P1-KAP1 D202-K434	KAP1 D202-K434 cloned into pGEX6P1	BamHI & EcoRI	Amp	ML
pGEX6P1-KAP1 L592-P835	KAP1 L592-P835 cloned into pGEX6P1	BamHI & EcoRI	Amp	ML
pGEX6P1-SUMO1GG	Human SUMO1 cDNA (terminates at di-glycine) cloned into pGEX6P1	BamHI & NotI	Amp	ML
pGEX6P1-SUMO2GG	Human SUMO2 cDNA (terminates at di-glycine) cloned into pGEX6P1	BamHI & EcoRI	Amp	ML
pGEX6P1-SUMO3GG	Human SUMO3 cDNA (terminates at di-glycine) cloned into pGEX6P1	BamHI & EcoRI	Amp	ML
pGIPZ-SMARCAD1 V2LHS-51716	shRNA targeting human SMARCAD1 (TAGTCTGGCCTACTCTATG) cloned into the pGIPZ lentiviral vector	Unknown	Amp	Dharmacon
pLKO shCHD5	shRNA targeting human CHD5 (CGTGTTCCTTTACTCCCTCTA) cloned into the pLKO lentiviral vector	Unknown	Amp	(Egan et al., 2013)
pLKO shCtrl	A non-targeting shRNA sequence cloned into the pLKO lentiviral vector	Unknown	Amp	(Egan et al., 2013)
pOG44	Expresses Flp recombinase for use with Flp-In system	N/A	Amp	Invitrogen
pUlp1	Recombinant Ulp1 (SUMO protease) bacterial expression vector	Unknown	Amp	PC
pVP35	VP35 cDNA with N-terminal HA tag cloned into pcDNA3.1(-)	Unknown	Amp	(Gantke et al., 2013)

Table 2.1 – Plasmids Used in this Study
In this table, thick borders demarcate groups of plasmids that share a common backbone. Several plasmids used in this study are commercially available, whereas others were generous gifts of others; the sources of all plasmids are clearly indicated. (DM = Daniel Maskell; HW = Hannah Williams; PC = Peter Cherepanov; PL = Paul Lesbats)

2.2.2 Polymerase Chain Reaction (PCR)

For cloning, PCR was performed with the high fidelity KOD DNA polymerase in accordance with the manufacturer's instructions (Toyobo). However, colony PCR reactions were performed using GoTaq® DNA polymerase (Promega). The primers were synthesized by Sigma-Aldrich. Each PCR reaction was optimized by including or excluding dimethyl sulfoxide (DMSO) in the reaction, and by varying the extension temperature using a gradient thermocycler.

2.2.3 Agarose Gel Electrophoresis

Gel electrophoresis was performed to resolve DNA by size using 0.8-2% (w/v) gels (depending on the expected size of the DNA fragments to be resolved), which contained 0.1µL/mL of SafeView (NBS Biologicals). 6X DNA loading buffer (***2.1.3.3***) was added to samples prior to electrophoresis. The gels were run in TBE at a constant voltage of 100V, typically for 30-45 minutes.

2.2.4 DNA Purification

DNA was purified from PCR reactions or following restriction digestion using the GeneJET™ PCR Purification kit (Thermo Fisher Scientific) as per the manufacturer's instructions. Likewise, DNA was recovered after agarose gel electrophoresis using the GeneJET™ Gel Extraction kit (Thermo Fisher Scientific).

2.2.5 Cloning

Cloning was performed according to conventional protocols by ligating together fragments of DNA that had been previously digested with restriction enzymes. Typical restriction digest reactions were performed in CutSmart® buffer with High-Fidelity® enzymes (New England Biolabs) for 3 hours at 37°C. Digested vectors were dephosphorylated by adding recombinant shrimp alkaline phosphatase (Roche) to the restriction digest reaction and incubating at 37°C for another hour. Digested DNA was then resolved by agarose gel electrophoresis and extracted

using the GeneJET™ Gel Extraction kit (Thermo Fisher Scientific). DNA ligase (Roche) was used to ligate the insert into the vector; the reaction was usually performed at room temperature for 1 hour, though it was occasionally incubated at 16°C overnight for convenience. Ligation reactions were transformed into chemically competent *E. coli* cells and colonies were assessed by colony PCR. The inserts of all cloned plasmids were sequenced to confirm their identities.

2.2.6 Site-Directed Mutagenesis

For site-directed mutagenesis, PCR reactions using primers that encoded for the point mutations were performed. The primers were designed in such a way that the region of dissimilarity was flanked by regions of complementarity (of at least 10bp in length each). The 5' ends of the forward and reverse primers abutted against each other, ensuring that the entire plasmid would be replicated, albeit in linear form. The PCR product was phosphorylated with T4 polynucleotide kinase (New England Biolabs), ligated into a circular plasmid using DNA ligase (Roche) and transformed into chemically competent *E. coli* cells.

2.2.7 Sequencing

Conventional Sanger sequencing was performed according to standard protocols by the Equipment Park core facility of the London Research Institute (LRI) latterly, The Francis Crick Institute.

2.3 Protein Techniques

2.3.1 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Criterion™ pre-cast 4-12% gradient polyacrylamide gels (Bio-Rad) gels were used to separate denatured proteins by size. Electrophoresis was conducted in Criterion™ gel tanks (Bio-Rad) in either MOPS or MES SDS-PAGE running buffer

(***2.1.3.8***), depending on the molecular weight range where maximal resolution was desired.

2.3.2 InstantBlue Staining

Following SDS-PAGE, polyacrylamide gels were immersed in InstantBlue™ solution (Expedion) and allowed to stain for at least 15 minutes, before being de-stained in water, as per the manufacturer's instructions. While InstantBlue™ stains proteins non-specifically, the staining intensity is nevertheless affected by the precise amino acid composition of a protein since the blue stain results from the Coomassie Brilliant Blue dye binding to basic amino acids (i.e. lysine, arginine and histidine) under acidic conditions.

2.3.3 Silver Staining

Silver staining of polyacrylamide gels was performed using the SilverQuest™ Silver Staining Kit (Invitrogen) according to the manufacturer's instructions. Silver staining depends on silver ions binding to specific functional groups; thus, strong staining is associated with amino acids containing carboxylic acid groups (i.e. aspartic acid and glutamic acid), imidazoles (i.e. histidine), sulfhydryls (i.e. cysteine) and amines (i.e. lysine). Hence, while silver staining is extremely sensitive, it is nonetheless affected by amino acid composition.

2.3.4 Western Blotting

After SDS-PAGE, proteins were transferred from polyacrylamide gels to nitrocellulose membranes (either Hybond C or Amersham™ Protran Premium 0.45µm nitrocellulose, GE Healthcare) using the Criterion™ Blotter apparatus (Bio-Rad). The transfer cassette was assembled in the following arrangement: sponge, three pieces of Whatman® filter paper, nitrocellulose membrane, polyacrylamide gel, three pieces of Whatman® filter paper and sponge. The tank was filled with Transfer Buffer (***2.1.3.12***) before performing electrophoresis at a constant wattage of 500mAmps for 90 minutes.

The membrane was stained with Ponceau S solution (Sigma) to confirm efficient protein transfer, and to assess the equality of loading. An image of the Ponceau S-stained membrane was acquired with a scanner. The membrane was then blocked in 5% (w/v) milk dissolved in PBS-Tween (***2.1.3.7***) for 30 minutes at room temperature. Proteins were detected by incubating the membranes with antibody solutions diluted in 5% (w/v) milk/PBS-Tween for either 1 hour at room temperature or overnight at 4°C. The antibody dilutions used in this study are listed in Table 2.2. Incubations were performed on a rocking platform to ensure that the antibody solution came into contact with the membrane in its entirety. Primary antibody solutions were reused several times; they were stored at 4°C between uses and were supplemented with 0.1% (w/v) sodium azide to prevent microbial contamination. The membrane was washed thrice in PBS-Tween, each time for 5 minutes, before being incubated with secondary antibody, diluted 1:10000 in 5% (w/v) milk/PBS-Tween, for 1 hour at room temperature. The secondary antibodies used were as follows: sheep anti-mouse IgG HRP-linked $F(ab')_2$ fragments and donkey anti-rabbit IgG HRP-linked $F(ab')_2$ fragments (GE Healthcare). Following incubation with secondary antibody, the membranes were again washed thrice in PBS-Tween in 5 minutes cycles.

To visualize the blots, the membrane was incubated with SuperSignal™ West Pico/Dura Chemiluminescent Substrate (Thermo Fisher Scientific) for 5 minutes. Luminescence was detected with Amersham™ Hyperfilm ECL (GE Healthcare). An Epson Perfection V700 Photo scanner was used to convert the data into a digital image, typically captured in 8-bit greyscale at a resolution of 600dpi.

For quantitative comparison of Western blotting data, the Odyssey® infrared fluorescence system (LI-COR) was used. Thus, instead of HRP-conjugated secondary antibodies, secondary antibodies conjugated to dyes emitting in the near-infrared range were used. Specifically, the donkey anti-mouse IgG IRDye® 680RD and goat anti-rabbit IgG IRDye® 800CW secondary antibodies (LI-COR) were used. Membranes were incubated with fluorescent secondary antibody solutions, diluted 1:5000 in 5% (w/v) milk/PBS-Tween for 1 hour at room temperature. After being washed thrice in PBS-Tween for 5 minutes each time, the

blots were visualized with an Odyssey® Infrared Imager (LI-COR). The signal from fluorescently labelled secondary antibodies is linear over a wider dynamic range as compared to standard Western blotting that relies upon a chemiluminescent enzymatic reaction.

Epitope	Clone/Product Number	Supplier	Antibody Type	Dilution
α-Tubulin	B-5-1-2	Sigma	Mouse monoclonal	1:3000
α-Tubulin	TAT-1	LRI	Mouse monoclonal	1:10000
CHD1	2F11H5	Millipore	Rat monoclonal	1:100
CHD5	ab66516	Abcam	Rabbit polyclonal	1:1000
FLAG	M2, F1804	Sigma	Mouse monoclonal	1:1000
FLAG	F7425	Sigma	Rabbit polyclonal	1:1000
Gata4	ab84593	Abcam	Rabbit polyclonal	1:1000
Gatad2a (p66α)	ab87663	Abcam	Rabbit polyclonal	1:1000
Glyr1	ab124615	Abcam	Rabbit polyclonal	1:1000
HA	12CA5	LRI	Mouse monoclonal	1:10000
HA	ab9110	Abcam	Rabbit polyclonal	1:4000
HDAC1	ab7028	Abcam	Rabbit polyclonal	1:1000
Histone H3	ab1791	Abcam	Rabbit polyclonal	1:1000
KAP1	4E1	Cell Signaling	Mouse monoclonal	1:2000
KAP1	ab10483	Abcam	Rabbit polyclonal	1:1000
KIAA1045	ab178954	Abcam	Rabbit polyclonal	1:500
MAP2	AP20, MAB3418	Millipore	Mouse monoclonal	1:1000
MTA1	ab71153	Abcam	Rabbit polyclonal	1:1000
MTA2	A300-395A	Bethyl	Rabbit polyclonal	1:1000
Nanog	ab80892	Abcam	Rabbit polyclonal	1:1000
NeuN	A60, MAB377	Millipore	Mouse monoclonal	1:1000
Oct4	ab19857	Abcam	Rabbit polyclonal	1:1000
RBBP7 (RbAp46)	ab3535	Abcam	Rabbit polyclonal	1:500
RPB1	4H8	LRI	Mouse monoclonal	1:20000
RPB1	8WG16	LRI	Mouse monoclonal	1:10000
Sbf1	ab177146	Abcam	Rabbit monoclonal	1:1000
SMARCAD1	A301-593A	Bethyl	Rabbit polyclonal	1:1000
Spindlin1	ab118784	Abcam	Rabbit polyclonal	1:1000
SUMO	N/A	Helle Ulrich	Rabbit polyclonal	1:10000
TIF1α	NB100-2596	Novus	Rabbit polyclonal	1:1000
Ubiquitin	P4D1	Enzo Life Sciences	Mouse monoclonal	1:1000

Table 2.2 – Antibodies Used for Western Blotting
The antibodies that were used in this study for Western blotting are listed here, along with the concentration at which they were used successfully.

2.3.5 Measurement of Protein Concentration

Protein concentrations were typically determined using the Bradford assay (Bio-Rad Protein Assay). A defined amount of the protein sample was added to 1mL of Bradford assay reagent in a disposable spectrophotometer cuvette; the reaction was allowed to proceed for 5 minutes at room temperature. The absorbance at 595nm was then measured with an Ultraspec™ 1100 pro spectrophotometer

(Amersham Biosciences). The protein concentration was then calculated based on a BSA standard curve. Two independent measurements were usually made for each sample and the average was adopted as the approximate protein concentration for the sample. The colorimetric change observed is the result of Coomassie Brilliant Blue G-250 dye binding to basic amino acid residues; thus, the protein concentration determined using the Bradford assay might be inaccurate, particularly for purified protein samples with unusual amino acid compositions.

As a complementary technique, the concentration of a purified protein sample was also assessed by visual comparison of the sample to defined quantities of BSA on an InstantBlue™-stained polyacrylamide gel. Similarly, the concentration of a purified protein sample was sometimes also determined by measuring the absorbance at 280nm using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific) and interpreting it with reference to the calculated extinction coefficient.

2.3.6 Immunoprecipitation (IP)

Cell extract containing 2.5mg of total protein was used for a typical immunoprecipitation reaction. As different samples usually had different protein concentrations, the reaction volumes were normalized with the addition of additional lysis buffer, usually to a final volume of 250-300µL. The IP reactions were frequently supplemented with benzonase/micrococcal nuclease and $MgCl_2$/$CaCl_2$ to final concentrations of 0.1U/µL and 2mM respectively; this step is designed to reduce proteins co-immunoprecipitating because of indirect interactions mediated by DNA. Primary antibody was then added to the extract and allowed to incubate at 4°C for 3 hours or overnight; typically, a ratio of 2µg of antibody to 1mg of total protein was used, though the efficiency with which individual antibodies immunoprecipitated their targets varied considerably. 15µL of Dynabeads® Protein A or G (Thermo Fisher Scientific), which had previously been equilibrated in the lysis buffer, was added to each IP reaction and allowed to incubate at 4°C for 1 hour. The unbound sample was then separated from the beads using a magnetic tube holder. The beads were then washed thrice; each wash step involved thorough resuspension of the beads in 500µL of lysis buffer,

before discarding the wash buffer with the aid of a magnetic tube holder. The beads were finally resuspended in 30µL of 2X SDS loading buffer and heated at 100°C for 5 minutes to elute all bound proteins.

FLAG-tagged proteins were immunoprecipitated with higher efficiency with the anti-FLAG® M2 affinity gel (Sigma) than with free anti-FLAG antibodies. In these cases, 15µL of pre-equilibrated M2 agarose beads per sample was added directly to the protein extract and allowed to incubate at 4°C for 3 hours. The beads were then subjected to a similarly rigorous series of wash steps; however, in this case, the liquid phase was separated from the beads using a Pierce™ spin column (Thermo Fisher Scientific).

2.4 Bacterial Techniques

2.4.1 Transformation of Chemically Competent *E. coli* Cells

One Shot® Top10 competent cells (Invitrogen) were typically used for cloning and for plasmid propagation. XL-10 Gold ultra-competent cells (Stratagene) were used for cloning large plasmids. Transformations of plasmids into these cells were performed as per the manufacturers' instructions.

For bacterial expression of recombinant proteins, plasmids were transformed into either BL21-CodonPlus® (DE3)-RIL or BL21-CodonPlus® (DE3)-RP competent cells (Stratagene), depending on the amino acid composition of the protein to be expressed. Transformed cells were selected against both the resistance marker of the plasmid and chloramphenicol to ensure the retention of additional copies of rare tRNA genes. Co-expression of recombinant proteins in *E. coli* was achieved by transforming cells with two plasmids simultaneously.

2.4.2 Purification of Plasmid DNA

Plasmid DNA was generally purified from 5mL of overnight *E. coli* cultures using the GeneJET™ Plasmid Miniprep kit (Thermo Fisher Scientific) as per the

manufacturer's instructions. Larger quantities of plasmid DNA were purified from 400mL of overnight bacterial cultures using the Plasmid Maxi kit (Qiagen).

2.4.3 Expression of Recombinant Proteins

The optimal conditions for bacterial expression of each recombinant protein were determined empirically in small-scale induction trials comparing the induction temperature and duration (Table 2.3). Nevertheless, a general protocol is described here. Starter cultures of transformed BL21-CodonPlus® (DE3)-RIL or BL21-CodonPlus® (DE3)-RP cells (Stratagene) were grown overnight at 37°C, in an incubator shaking at 180rpm, in LB (***2.1.1.1***) supplemented with the appropriate antibiotic for that plasmid and chloramphenicol. 1L cultures of LB were inoculated with 12mL of the overnight starter culture and cultured at 37°C, for approximately 2-3 hours, until an optical density (OD_{600}) of 0.6-0.8 was reached, to ensure that the cells were in mid-log phase. LB was supplemented with 50μM zinc sulphate for the expression of proteins with zinc fingers. Recombinant protein expression was induced by the additional of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5mM. The cultures were then shifted to either 30°C or 16°C and cultured for an additional 3 or 6 hours respectively. The cells were harvested by centrifugation in a chilled centrifuge at 6000×g for 30 minutes, washed in PBS (***2.1.3.6***), and snap-frozen in liquid nitrogen. The *E. coli* pellets were stored at -80°C until they were used for protein purification. In some cases, 30L cultures were grown in a fermenter with the assistance of the Fermentation core facility (LRI).

Protein	**Plasmid**	***E. coli* Strain**	**Induction Conditions**		
			[IPTG]	**Temperature**	**Duration**
SMARCAD1	pET28a-SUMO-SMARCAD1 & tagged derivatives	RIL	0.5mM	16°C	6 hours
CUE1,2 (S95-N347)	pET28a-SUMO-CUE1,2 & tagged derivatives	RIL	0.5mM	30°C	3 hours
KAP1	pET28a-SUMO-KAP1 & tagged derivatives	RP	0.5mM	30°C	6 hours
KAP1 RBCC (S33-K434)	pET28a-SUMO-KAP1 S33-K434	RP	0.5mM	30°C	6 hours
SMARCAD1 S95-N347 + KAP1 S33-K434	pET28a-SUMO-CUE1,2 (S95-N347) + pCDF-SUMO-KAP1 S33-K434	RP	0.5mM	30°C	6 hours
Ulp1	pUlp1	RIL	0.1mM	30°C	4 hours
Ubiquitin	pGEX-Ub & derivatives; pET42a-Ub + derivatives	RP	1mM	30°C	3 hours

Table 2.3 – Conditions Used for Bacterial Expression of Recombinant Proteins
The conditions in this study used to express recombinant proteins in E. coli are listed in this table. Most of these conditions have been optimized empirically in small-scale induction trials comparing several induction temperatures and durations. In terms of *E. coli* strains, RIL and RP refer to BL21-CodonPlus(DE3)-RIL and BL21-CodonPlus(DE3)-RP cells (Stratagene) respectively.

2.4.4 Preparation of Bacterial Whole Cell Lysates

1mL of an *E. coli* culture was pelleted by centrifugation at 4000×g for 10 minutes. The supernatant was discarded and the pellet was washed once in 500µL of PBS. The cell pellet was then resuspended in 200µL of 1X SDS-PAGE loading buffer (***2.1.3.1***). The release of DNA from the cells results in a significant increase in sample viscosity. Thus, the sample was sonicated with a tip sonicator (Branson) for 10 seconds at an amplitude of 10%, or until the viscosity of the sample was significantly reduced. The sample was then centrifuged at 14000rpm in a benchtop centrifuge (Eppendorf) for 10 minutes to obtain a denatured, bacterial whole cell lysate.

2.5 Techniques for Mammalian Cells

2.5.1 Standard Culturing of Mammalian Cells

Cells were usually cultured in DMEM (***2.1.2.1***) as adherent monolayers in a 37°C incubator with 3% CO_2 according to standard protocols. Cultures were split upon reaching approximately 80% confluency – cells were dissociated from the plate by brief incubation with a trypsin/EDTA solution (***2.1.2.5***) and a proportion of the cells was re-plated. To freeze cells, they were resuspended in freezing medium (***2.1.2.3***), transferred into cryogenic vials, and frozen using a freezing container that achieves a rate of cooling of -1°C/minute.

2.5.2 Preparation of Irradiated MEF Feeder Cells

Irradiated DR-4 MEFs were prepared by the Cell Services core facility (LRI). 12.5-13.5 days post-coital mouse embryos were dissected from gravid females and washed in sterile PBS (***2.1.3.6***). The embryos were then cut into small pieces using

a sterile pair of scissors. The sample was then digested with a trypsin/EDTA (***2.1.2.5***) solution at 37°C until there were no more large pieces; this process usually took approximately 30 minutes, at which point foetal bovine serum was added to inhibit trypsin. The cells were then plated on 10cm plates – generally, each embryo was sufficient for 2 × 10cm plates – and cultured in DMEM (***2.1.2.1***) until confluency, upon which they were frozen (**2.5.1**). To support the growth of mouse ESCs as a layer of feeder cells, 3×10^6 MEFs were plated on a 10cm plate and mitotically inactivated with 4Gy of γ-irradiation.

2.5.3 Culturing of Mouse ESCs

Mouse ESCs were cultured on feeder layers of irradiated MEFs (**2.5.2**) in mouse ESC medium (***2.1.2.2***) in a 37°C incubator with 3% CO_2. Due to the rapid proliferation rate of these cells, cultures were split (**2.5.1**) every other day. The culture medium was refreshed on days between splits.

2.5.4 Differentiation of Mouse ESCs into Embryoid Bodies

Mouse ESCs were induced to differentiate by switching them to suspension culture in the absence of leukaemia inhibitory factor (LIF). Mouse ESCs were dissociated from plates by incubation with a trypsin/EDTA solution (***2.1.2.5***). Re-plating the cells on fresh cell culture plates and collecting the culture medium 30 minutes later recovered a sample of ESCs largely depleted of MEFs, which adhere more readily to the plate. 1×10^7 ESCs were then transferred to each 10cm bacterial petri dish and cultured in ESC medium lacking LIF. The culture medium was refreshed every two days. As embryoid bodies are relatively delicate structures whose integrity can be compromised by centrifugation, the culture was transferred to a 15mL Falcon tube and incubated for at least 5 minutes to allow the embryoid bodies to sediment by gravity, permitting the medium to be removed using a pipette. Embryoid bodies were cultured for up to seven days; over that time period, they gradually transformed from cell aggregates into large cystic structures.

2.5.5 Neuronal Differentiation of SH-SY5Y Neuroblastoma Cells

1×10^5 SH-SY5Y cells were seeded in each well of a 6-well plate coated with poly-D-lysine. (To coat the plates, each well was incubated with 1.5mL of a 50mg/mL poly-D-lysine solution for 1 hour at room temperature, washed with 1.5mL of water twice, and allowed to dry for 1 hour.) The next day, the medium was replaced by DMEM (***2.1.2.1***) with only 0.5% (v/v) foetal bovine serum, but supplemented with 1X N-2 Supplement (Gibco, Thermo Fisher Scientific) and 5μM retinoic acid (Sigma). As retinoic acid is a photosensitive compound, the plates were wrapped in aluminium foil. The cells were allowed to differentiate for as long as 12 days, but features of neuronal differentiation could be observed 3 days after initiating differentiation.

2.5.6 Calcium Phosphate Transfection

For calcium phosphate transfection of a 15cm plate of 293T cells, 2.5mL of a solution containing 250mM $CaCl_2$ and plasmids was prepared. The amount of each plasmid added varied; for example, optimal overexpression of CHD1 required 36μg of pCHD1-FTH/FRT and 18μg of pVP35, while CHD5 was overexpressed with 36μg of pCHD5-FTH/FRT and 9μg of pVP35. The $CaCl_2$-plasmid mixture was added to 2.5mL of 2X BES buffered saline (Sigma) in a drop-wise manner. While adding the $CaCl_2$-plasmid mixture, bubbles were generated in the BES solution to promote the formation of calcium phosphate crystals. The mixture was then incubated at room temperature for 15 minutes, before it was added to a plate of 293T cells that had been plated the day before. The culture medium was replaced the next morning to limit cell toxicity. The transfected cells were harvested 24 hours after transfection. This protocol can be scaled up or down depending on the number of plates and the size of plates to be transfected.

2.5.7 Lipofectamine® 2000 Transfection

Transfection with Lipofectamine® 2000 (Invitrogen, Thermo Fisher Scientific) was performed in accordance with the manufacturer's instructions. For example, to transfect a 10cm plate of 293T cells, a total of 20μg o fplasmid DNA was diluted in

500µL of Opti-MEM® medium (***2.1.2.4***). Likewise, 70µL of Lipofectamine® 2000 lipid reagent was diluted with 430µL of Opti-MEM® medium. The diluted DNA was added to the diluted lipid transfection reagent and the mixture was incubated at room temperature for 5 minutes. The mixture was then added to a 70-80% confluent plate of 293T cells. The culture medium was replaced the next day to minimize cell toxicity.

2.5.8 Generation of Stable Cell Lines

The mouse ESCs expressing tagged CHD proteins used in this study (Table 2.4) were generated by homologous recombination by our collaborator, Matthieu Gérard (IBITECS CEA Saclay, France).

The Flp-In™ system (Invitrogen, Thermo Fisher Scientific) was used to generate 293 cell lines (Table 2.4) that stably overexpressed specific proteins. Typically, a 10cm plate of 293 Flp-In™ cells (i.e. 293 cells that contain FRT sites) were transfected with 18µg of pOG44, which encodes for the Flp-recombinase, and 2µg of a pcDNA5/FRT-derived plasmid encoding for the desired protein using the Lipofectamine® 2000 transfection method (**2.5.7**). The cells were subsequently selected for hygromycin resistance. Individual colonies were isolated using a cloning cylinder, detached from the plate using a trypsin/EDTA solution (***2.1.2.5***), and re-plated in a 48-well plate. The 293 Flp-In™ cells stably expressing either wild type SMARCAD1 or a version with point mutations in both CUE domains (F169A, P170A, L195A, L196A, F263A, P264A, L289A, K290A) were generated by Hannah Williams.

A 293 T-REx™ cell line (Invitrogen, Thermo Fisher Scientific) that was depleted of endogenous SMARCAD1 by shRNA knockdown, but expressed exogenous FLAG-tagged SMARCAD1 from a doxycycline inducible promoter was also established by Hannah Williams. This was achieved by transfecting 293 T-Rex cells with pSMARCAD1 716R/TO using Lipofectamine® 2000 (**2.5.7**) and selecting for Zeocin™ (Thermo Fisher Scientific) resistance. The pool of Zeocin™-resistant cells were then transduced with the V2LHS_51716 shRNA (Dharmacon) using the

pGIPZ™ lentiviral vector (Dharmacon), before being selected for puromycin resistance. Individual colonies were isolated as described above. A similar cell line that rescued the cells with exogenous SMARCAD1 bearing point mutations in both CUE domains was also generated. By titrating the doxycycline concentration, it was determined by quantitative Western blotting using the Odyssey® infrared fluorescence system (**2.3.4**) that exogenous SMARCAD1 was expressed at approximately endogenous level in the presence of 1ng/mL and 0.5ng/mL doxycycline in the 293 TREX F-SMARCAD1 716R and 293 TREX F-SMARCAD1 CUE1mt,2mt 716R cell lines respectively.

Cell Line	Description	Source
CHD1-FTH mESCs	Endogenous *CHD1* locus in 46C mESCs engineered to contain C-terminal FLAG-TEV-HA (tag	MG
CHD2-FTH mESCs	Endogenous *CHD2* locus in 46C mESCs engineered to contain C-terminal FLAG-TEV-HA tag	MG
CHD3-FTH mESCs	Endogenous *CHD3* locus in 46C mESCs engineered to contain C-terminal FLAG-TEV-HA tag	MG
CHD4-FTH mESCs	Endogenous *CHD4* locus in 46C mESCs engineered to contain C-terminal FLAG-TEV-HA tag	MG
CHD5-FTH mESCs	Endogenous *CHD5* locus in 46C mESCs engineered to contain C-terminal FLAG-TEV-HA tag	MG
CHD6-FTH mESCs	Endogenous *CHD6* locus in 46C mESCs engineered to contain C-terminal FLAG-TEV-HA tag	MG
CHD8-FTH mESCs	Endogenous *CHD8* locus in 46C mESCs engineered to contain C-terminal FLAG-TEV-HA tag	MG
CHD9-FTH mESCs	Endogenous *CHD9* locus in 46C mESCs engineered to contain C-terminal FLAG-TEV-HA tag	MG
293 Flp-In F-SMARCAD1 C3	293 Flp-In cells stably overexpressing FLAG-tagged SMARCAD1; highest level of overexpression	HW
293 Flp-In F-SMARCAD1 C5	293 Flp-In cells stably overexpressing FLAG-tagged SMARCAD1	HW
293 Flp-In F-SMARCAD1 CUE1mt,2mt C2	293 Flp-In cells stably overexpressing FLAG-tagged SMARCAD1 that has point mutations (F169A, P170A, L195A, L196A, F263A, P264A, L289A, K290A) in both CUE domains; much lower expression that wild type cell lines	HW
293 Flp-In F-SMARCAD1 CUE1mt,2mt C7	293 Flp-In cells stably overexpressing FLAG-tagged SMARCAD1 that has point mutations F169A, P170A, L195A, L196A, F263A, P264A, L289A, K290A) in both CUE domains; lowest level of overexpression	HW
293 Flp-In HA-KAP1	293 Flp-In cells stably overexpressing HA-tagged KAP1	ML
293 Flp-In HA-KAP1 10×Ub-mt	293 Flp-In cells stably overexpressing HA-tagged KAP1 with K → R point mutations at 10 ubiquitylation sites	ML
293 T-REx F-SMARCAD1 716R	293 TREX cells depleted of endogenous SMARCAD1 (due to stable knockdown by the V2LHS_51716 shRNA) that expresses FLAG-tagged SMARCAD1 from a doxycycline-inducible promoter	HW
293 T-REx F-SMARCAD1 664R	293 TREX cells depleted of endogenous SMARCAD1 (due to stable knockdown by the V3LHS_343664 shRNA) that expresses FLAG-tagged SMARCAD1 from a doxycycline-inducible promoter	HW
293 T-REx F-SMARCAD1 CUE1mt,2mt 716R	293 TREX cells depleted of endogenous SMARCAD1 (due to stable knockdown by the V2LHS_51716 shRNA) that expresses FLAG-tagged SMARCAD1 with point mutations in both CUE domains from a doxycycline-	ML

	inducible promoter	

Table 2.4 – Cell Lines Generated or Used in this Study
The cell lines used in this study are described in this table. Many of these cell lines were generous gifts of collaborators, most notably Matthieu Gérard (MG) and Hannah Williams (HW); the provenance of each cell line is clearly indicated.

2.5.9 Preparation of Whole Cell Extracts

Cell pellets were thoroughly resuspended in Triton lysis buffer (***2.1.3.13***) supplemented with 1X protease inhibitor cocktail (***2.1.3.4***) and allowed to incubated on ice for 30 minutes; a typical pellet from a 10cm plate was resuspended in 120µL of lysis buffer. The samples were then sonicated in a cold water bath using the Bioruptor® (Diagenode) for 3-5 minutes in 30s on/30s off cycles at high intensity. Some proteins, including CHD5, were adversely affected by sonication; hence when performing a Western blot for CHD5, the samples were not sonicated. The samples were then clarified by centrifugation at 14000rpm for 20 minutes at 4°C in an Eppendorf 5417R microcentrifuge.

2.6 Protein Purification from Mammalian Tissues

2.6.1 Preparation of Chromatin Extracts with Heparin

Each mouse ESC pellet was resuspended in four times its volume of WCE-150 buffer (***2.1.4.1***). Extensive cell lysis was achieved by processing the cell resuspension with loose and tight Dounce homogenizers, and then by passing it through a 21G needle thrice. The sample was then centrifuged at 45000rpm in a Type 45 Ti rotor (Beckman Coulter) for 1 hour. The supernatant was discarded, while the pellet was resuspended in B-100 buffer (***2.1.4.2***), using the same volume of B-100 buffer as WCE-150 buffer used initially. The pellet was effectively resuspended by sonication of the sample for 1.5 minutes in 15s on/30s off cycles at 25% amplitude. Heparin was mixed into the sample, which was then incubated at 4°C for 10 minutes. The amount of heparin added was based on an approximate DNA concentration of 2.5mg/mL extract. As extensive extraction of chromatin-bound proteins (including CHD1) was observed with 2.5 times (w/w) the amount of heparin to DNA, this was the ratio typically used. Benzonase® (Millipore), RNase A and $MgCl_2$ were then added to the sample to final concentrations of 200U/mL,

2mg/mL and 2mM respectively. The sample was incubated for 30 minutes before being spun at 45000rpm in a type 45 Ti rotor (Beckman Coulter). The supernatant of the second spin was the soluble chromatin extract.

2.6.2 Purification of RNAPII from Mouse ESCs

Approximately 4mL of chromatin extract (**2.6.1**) was mixed with 200µL of pre-equilibrated 8WG16 beads – an equal mixture of Pierce™ Protein A and G agarose beads (Thermo Fisher Scientific) that was covalently coupled with the mouse monoclonal antibody, 8WG16, using the dimethyl pimelimidate technique – and incubated at 4°C for 3 hours. As a control, mouse IgG beads were similarly incubated with an equal amount of chromatin extract. To enhance the binding efficiency, the unbound samples from the initial batch-binding process were pumped over the resins at a flow rate of 1mL/minute.

For the low stringency purification, the beads were washed twice, using 35mL of R-150N buffer (***2.1.4.3***) supplemented with 0.01% (v/v) NP-40 per wash. Bound RNAPII was then eluted from the resin by incubation at 4°C for 30min with 200µL of 1mg/mL CTD peptide (Peptide Chemistry core facility, LRI) dissolved in RG-150N buffer (***2.1.4.3***). This elution step was repeated 5 times in total; the final cycle was performed at 37°C to maximise protein elution. The high stringency purification was performed in a similar manner; however, the washes were performed with R-500Am buffer (***2.1.4.4***) supplemented with 0.1% (v/v) NP-40, while the CTD peptide was dissolved in RG-500Am (***2.1.4.4***) buffer for elution.

2.6.3 Purification of CHD1 from Mouse ESCs

60mL of chromatin extract (**2.6.1**) from (90 × 15cm plates of) mouse ESCs expressing C-terminally FLAG-tagged CHD1 was mixed with 500µL of pre-equilibrated anti-FLAG® M2 affinity gel (Sigma) and incubated at 4°C for 3 hours. As a control, an equal amount of chromatin extract from parental 46C mouse ESCs was used for a parallel purification performed identically. The unbound samples were re-loaded over the resin by pumping at a flow rate of 1mL/minute. For the low

stringency purification, the columns were washed in 20mL (40CV) of CHD-150 (***2.1.4.5***) buffer supplemented with 0.01% (v/v) NP-40 thrice. CHD1 was eluted from the resin by incubation for 30min at 4°C with 500µL of 1mg/mL FLAG peptide (Peptide Chemistry core facility, LRI) dissolved in CHD-150 buffer. This elution step was repeated 5 times in total; the final cycle was performed at 37°C to maximise protein elution. The fractions containing CHD1 were pooled, mixed with 60µL of pre-equilibrated anti-HA (3F10) affinity matrix (Roche) and allowed to incubate at 4°C for 3 hours. The HA beads were washed in 1mL of CHD-150 buffer twice, before being equilibrated in TEV buffer (***2.1.4.8***). CHD1 was eluted from the HA resin by cleavage with TEV protease – the beads were incubated with 60µL of 200µg/mL TEV protease diluted in TEV buffer for 1.5 hours at room temperature. The high stringency purification was performed in a comparable manner though different wash buffers were used – the beads were washed once with CHD-750 buffer (***2.1.4.6***) supplemented with 0.01% (v/v) NP-40 and once with CHD-50 buffer (***2.1.4.7***) supplemented with 0.01% (v/v) NP-40.

2.6.4 Purification of CHD5 Overexpressed in 293T Cells

10 × 15cm plates of 293T cells were transfected using calcium phosphate with the pCHD5-FTH/FRT and pVP35 plasmids (**2.5.6**), resulting in transient overexpression of CHD5. As a control, CHD1 was transiently overexpressed in 293T cells in parallel. The cell pellets were resuspended in 4X their volumes of CL-50 buffer (***2.1.4.9***) supplemented with 1X protease inhibitor cocktail (***2.1.3.4***). The samples were sonicated using the tip sonicator (Branson) for 3 minutes in 15s on/30s off cycles at 30% amplitude, with the samples being kept cool in an ice bath of salt and ice during the process. Following addition of Benzonase® (Millipore) to a concentration of 0.1U/mL, the sample was incubated at 4°C for 30 minutes. The sodium chloride concentration of the sample was then raised to 750mM to extract proteins from chromatin. A soluble extract was then obtained by centrifugation at 41000rpm in the SW 41 Ti rotor (Beckman Coulter) for 30 minutes at 4°C.

To purify the CHD5 complex, the CHD5 extract was loaded onto a 250µL column of pre-equilibrated anti-FLAG® M2 affinity gel (Sigma) at a flow rate of 0.1mL/minute.

As a control, CHD1 was purified in parallel. The columns were washed with CHD-750 (***2.1.4.6***), CHD-50 (***2.1.4.7***) and CHD-150 (***2.1.4.5***) buffers; for each wash, 12.5mL (50CV) of buffer supplemented with 0.01% (v/v) NP-40 was pumped over the column at a flow rate of 2mL/minute. Elution was performed as previously described (**2.6.3**), though each elution cycle used 250µL of 1mg/mL FLAG peptide (Peptide Chemistry core facility, LRI) dissolved in CHD-150 buffer.

2.6.5 Purification of CHD5 from Mouse Brains

17 frozen *CHD5*$^{3FTH/3FTH}$ mouse brains were ground into a fine powder using a ceramic mortar and pestle; to ensure that the sample remained frozen, the mortar was filled with liquid nitrogen during this process. The powdered brain sample was resuspended in C-100Ca buffer (***2.1.4.10***) supplemented with 1X protease inhibitor cocktail (***2.1.3.4***); 6µL of buffer was added per 1mg of mouse brains. 2000 gel units of micrococcal nuclease (NEB) was then added per mouse brain used; to promote digestion, the sample was incubated at room temperature for 30 minutes, while being mixed with a magnetic stir bar. The sodium chloride concentration of the sample was raised to 750mM, before micrococcal nuclease activity was inactivated by addition of EGTA to a concentration of 5mM. Soluble extract was obtained by centrifugation of the sample at 45000rpm in a Type 45 Ti rotor (Beckman Coulter) at 4°C for 30 minutes.

The mouse brain extract was pumped over a 250µL column of pre-equilibrated anti-FLAG® M2 affinity gel (Sigma) at a flow rate of 0.25mL/minute. The resin was washed with C-750 (***2.1.4.11***), C-50 (***2.1.4.12***) and C-100 (***2.1.4.10***) buffers; for each wash, 25mL (100CV) of buffer was pumped over the column at a flow rate of 2mL/minute. CHD5 was eluted from the column by incubation with 300µL of 500µg/mL FLAG peptide (Peptide Chemistry core facility, LRI) dissolved in C-100 buffer for 30 minutes at 4°C. To maximise the elution, it was repeated again, though this time the sample was incubated at 37C for 30 minutes.

The peak fraction was loaded onto an equilibrated Mini S™ PC 3.2/3 column (240µL) at a flow rate of 0.05mL/minute. The column was washed in 2CV of C-100

buffer, and CHD5 was eluted in a 4.8mL (20CV) salt gradient of 0-50% C-2000 (***2.1.4.13***) at a flow rate of 0.2mL/minute. Alternatively, CHD5 could also be purified using the MAbPac SCX-10G column (Thermo Fisher Scientific). CHD5 was eluted from this column in a 4.8mL salt gradient of 0-100% C-2000 at a flow rate of 0.2mL/minute.

2.6.6 Tandem Liquid Chromatography-Mass Spectromety (LC-MS) Analysis

LC-MS analysis was performed by the Protein Analysis and Proteomics core facility (LRI). Protein sample were submitted as gel slices, following which in-gel tryptic digestion and peptide extraction were performed according to standard protocols. Peptide mixtures were resuspended in 0.1% (v/v) TFA before separation by liquid chromatography. Hydrophilic protonated peptides were first concentrated on a trapping column before being separated on a 500mm × 75μm internal diameter Acclaim® Pepmap™ column (Thermo Fisher Scientific) and eluted directly onto an Orbitrap™ mass spectrometer (Thermo Fisher Scientific). Data-dependent acquisition was performed – the top 10 or 20 most abundant peptides were selected for MS/MS. Data analysis was subsequently performed using the MaxQuant bioinformatics suite.

2.7 Purification of Bacterially-Expressed Recombinant Proteins

2.7.1 Purification of SMARCAD1

Expression of N-terminally FLAG-tagged SMARCAD1 was induced as previously described (**2.4.3**) in a 10L culture of *E. coli* transformed with the pET28a-SUMO-F-SMARCAD1 plasmid or its derivatives. (While C-terminally FLAG-tagged SMARCAD1 is soluble, the C-terminal tag is inaccessible to M2 agarose.) The resulting cell pellet was resuspended in 60mL of HL-500 buffer (***2.1.5.1***) supplemented with 1X Protease Inhibitor cocktail (***2.1.3.4***). Thorough resuspension of such a large cell pellet by pipetting alone was frequently challenging; however, it was greatly assisted by the use of a loose Dounce homogenizer. The sample was

divided into three equal portions of approximately 30mL each; each portion sonicated using a tip sonicator with a large probe (Branson) for 2 minutes in 15s on/30s off cycles at 25% amplitude. During sonication, the samples were kept cool in an ice bath of salt and ice. The sonicated sample was subsequently spun at 35,000 rpm in a Type 45 Ti rotor (Beckman Coulter) for 30 minutes.

The supernatant was loaded through a 150mL Superloop™ (GE Healthcare) onto a 5mL HisTrap™ HP column (GE Healthcare) at a flow rate of 1mL/minute. The column was washed in 10 column volumes (CV) of HW-750 buffer (***2.1.5.2***) at a flow rate of 2mL/minute. The column was then washed in HW-100 (***2.1.5.3***) until the conductivity measurement stabilized at the baseline. Elution was started with a 30CV gradient from 0-30% of HE-300 buffer (***2.1.5.4***) at a flow rate of 1mL/minute and continued using 100% HE-300 until the A280 trace stabilized at the baseline; the elution was collected in 1mL fractions. SMARCAD1 tended to elute in 50mL across the entire gradient; however, limited resolution from contaminants was occasionally observed at the start of the elution, hence the use of an initial imidazole gradient. Fractions containing SMARCAD1 were pooled along with 5µL of Ulp1 (22mg/mL), before being dialyzed (12.4K MWCO) overnight at 4°C against 2L of P-100 buffer (***2.1.5.6***).

The dialyzed sample was loaded through a 50mL Superloop™ (GE Healthcare) onto a 5mL HiTrap™ Heparin HP column (GE Healthcare) at a flow rate of 1mL/minute. A first cut with 25% P-1000 buffer at a flow rate of 1.5mL/minute until the A280 reading stabilized at the baseline removed a dominant contaminant roughly 55kDa in size. Full-length SMARCAD1 was eluted in a 10CV gradient from 25-45% P-1000 buffer (***2.1.5.7***) at a flow rate of 0.8mL/minute; 1mL fractions were collected. The fractions were pooled conservatively since separation of the ~100kDa contaminant from SMARCAD1 from full-length SMARCAD1 was inefficient. The pooled sample was dialyzed overnight at 4°C against 2L of Q-100 buffer (***2.1.5.8***) for wild-type SMARCAD1 and the CUE1mt,2 mutant, and against 2L of P-50 buffer (***2.1.5.5***) for the CUE1,2mt and CUE1mt,2mt mutants.

While wild-type SMARCAD1 does not bind to ProSwift WCX-1S (Thermo Fisher Scientific), a weak cation exchange column, this is useful chromatographic step for

SMARCAD1 bearing point mutations in its second CUE domain, since full-length protein can be resolved from a ~100kDa contaminant. If applicable, sample was loaded through a 10mL Superloop™ (GE Healthcare) onto the ProSwift WCX-1S column at a flow rate of 0.5mL/minute. A 19mL elution gradient from 0-30% P-1000 buffer at 0.5mL/minute separated full-length protein from the ~100kDa contaminant; 500µL fractions were collected. Following conservative pooling, the sample was diluted in P-50 for a final NaCl concentration of approximately 100mM.

The final column was ProSwift SAX-1S (Thermo Fisher Scientific), a strong anion exchange column, to which samples were loaded via a 10mL Superloop™ at a flow rate of 0.5mL/minute. A sharp, initial 2.5mL gradient of 0-10% Q-1000 buffer (***2.1.5.9***) helped removed a few contaminants; however, SMARCAD1 eluted within a 17.5mL gradient of 10-40% Q-1000 buffer. Pooled fractions were concentrated in and exchanged into P-100 buffer using a Microcon® spin concentrator (Millipore) with a 50K molecular weight cut-off (MWCO).

2.7.2 Purification of SMARCAD1 CUE1,2 Fragment (S95-N347)

The cell pellet from a 1L culture of *E. coli* expressing the CUE1,2 fragment (**2.4.3**) was resuspended in 20mL of HL-500 buffer (***2.1.5.1***) supplemented with 1X Protease Inhibitor cocktail (***2.1.3.4***). The sample was sonicated for 3 minutes in 15s on/30s off cycles, while being kept cool in an ice bath of salt and ice. To obtain a soluble extract, the sample was spun at 50000rpm for 30 minutes in the Type 70 Ti rotor (Beckman Coulter). The sample was loaded onto 3mL of pre-equilibrated Ni-NTA agarose (Qiagen) at a flow rate of 0.5mL/minute. The column was first washed in 10CV of HW-750 buffer (***2.1.5.2***), then in 10CV of HW-100 (***2.1.5.3***) at a flow rate of 2mL/minute. The CUE domain fragment was eluted from the column with HE-300 buffer (***2.1.5.4***) into a sample volume of approximately 9mL, to which 2.5µL of Ulp1 (56mg/mL) was added. The sample was dialyzed (12.4K MWCO) against 2L of Q-100 buffer (***2.1.5.8***) for 1.5 hours at 4°C, before the buffer was replaced with another 2L of Q-100 buffer and dialysis allowed to continue for another 1.5 hours. To deplete the SUMO tag, the dialyzed sample was loaded back

over the 3mL of Ni-NTA resin at a flow rate of 0.5mL/minute and the unbound flow-through collected.

The sample was then loaded using a 10mL Superloop™ (GE Healthcare) onto an equilibrated 1mL Mono Q® 5/50 GL column (GE Healthcare) at a flow rate of 0.8mL/minute. The CUE domain fragment was eluted from the column within a 30mL salt gradient of 0-30% Q-1000 (***2.1.5.9***), run at a flow rate of 0.8mL/min, and collected in 500µL fractions. The peak fractions were pooled and dialyzed overnight against 2L of Q-100 buffer at 4°C, before being concentrated into an approximate total volume of 600µL using an Amicon® Ultra-4 10K MWCO spin concentrators (Millipore). The protein concentration of the final sample was in the 5mg/mL range.

2.7.3 Purification of KAP1

Expression of KAP1 was induced as previously described (**2.4.3**) in an 8L culture of *E. coli* transformed with the pET28a-SUMO-KAP1 plasmid or a tagged version. The resulting cell pellet was resuspended in 60mL of HL-500Zn buffer (***2.1.5.1***) supplemented with 1X Protease Inhibitor cocktail (***2.1.3.4***). Soluble extract was prepared by sonication and centrifugation as previously described for the SMARCAD1 purification (**2.7.1**). The sample was loaded through a 150mL Superloop™ onto a 5mL HisTrap™ HP column at a flow rate of 1mL/minute. The column was washed in 10CV of HW-750Zn buffer (***2.1.5.2***) at a flow rate of 2mL/minute. The column was then washed in HW-100Zn (***2.1.5.3***) until the conductivity measurement stabilized at the baseline. KAP1 was eluted from the column in 1mL fractions, using a 100mL imidazole gradient from 0-50% HE-300 buffer (***2.1.5.4***) at a flow rate of 1mL/minute. Fractions containing KAP1 were pooled; following addition of 5µL of Ulp1 (22mg/mL), the sample was dialyzed (12.4K MWCO) overnight at 4°C against 2L of P-100Zn buffer.

The dialyzed sample was loaded through a 50mL Superloop™ (GE Healthcare) onto a 5mL HiTrap™ Heparin HP column (GE Healthcare) at a flow rate of 1mL/minute. KAP1 was eluted from the column with a 100mL salt gradient from 0-50% P-1000Zn buffer (***2.1.5.7***), in 1mL fractions. Those containing KAP1 were

pooled and concentrated to a volume of approximately 4mL using an Amicon® Ultra-15 30K MWCO spin concentrator (Millipore).

The sample was then loaded using a 10mL Superloop™ onto a 120mL HiPrep™ 16/60 Sephacryl® S-400 HR gel filtration column (GE Healthcare) at a flow rate of 0.3mL/minute. The column was eluted isocratically with GF-150Zn buffer (***2.1.5.12***) over 1.1CV at a flow rate of 0.3mL/minute in 1.2mL fractions. KAP1 eluted as a single peak, which was pooled and concentrated to approximately 1.4mL (i.e. 9mg/mL) using an Amicon® Ultra-15 30K MWCO spin concentrator (Millipore).

2.7.4 Purification of KAP1 RBCC S33-K434

The cell pellet from a 1L culture of *E. coli* expressing the KAP1 S33-K434 (**2.4.3**) was resuspended in 20mL of HL-500Zn buffer (***2.1.5.1***) supplemented with 1X Protease Inhibitor cocktail (***2.1.3.4***). A soluble extract was prepared by sonication and centrifugation, as previously described for the CUE domain fragment (**2.7.2**). The sample was loaded onto 3mL of pre-equilibrated Ni-NTA agarose (Qiagen) at a flow rate of 0.25mL/minute. The column was first washed in 10CV of HW-750Zn buffer (***2.1.5.2***), then in 10CV of HW-100Zn (***2.1.5.3***) at a flow rate of 2mL/minute. KAP1 S33-K434 was eluted from the column with HE-300 buffer (***2.1.5.4***) into a sample volume of 9mL, to which 2.5µL of Ulp1 (56mg/mL) was added. The sample was dialyzed (12.4K MWCO) against 2L of Q-100Zn buffer (***2.1.5.8***) for 1.5 hours at 4°C, before the buffer was replaced with another 2L of Q-100Zn buffer and dialysis allowed to continue for another 1.5 hours. To deplete the SUMO tag, the dialyzed sample was loaded back over the 3mL of Ni-NTA resin at a flow rate of 0.25mL/minute and the unbound flow-through collected.

The sample was then loaded using a 10mL Superloop™ (GE Healthcare) onto an equilibrated 1mL Mono Q® 5/50 GL column (GE Healthcare) at a flow rate of 0.6mL/minute. The CUE domain fragment was eluted from the column within a 50mL salt gradient of 0-50% Q-1000Zn (***2.1.5.9***), run at a flow rate of 0.5mL/min, and collected in 500µL fractions. The peak fractions were pooled and dialyzed overnight against 2L of Q-100 buffer at 4°C. The sample was not concentrated any

further as it sticks non-specifically to Amicon® Ultra-4 spin concentrators (Millipore); nevertheless, the concentration of the sample was 0.4mg/mL.

2.7.5 Purification of SMARCAD1 CUE1,2 Fragment (S95-N347) in Complex with KAP1 S33-K434

The cell pellet from an 8L culture of *E. coli* expressing SMARCAD1 S95-N347 and KAP1 S33-K434 (**2.4.3**) was resuspended in 50mL of HL-200Zn buffer (***2.1.5.10***) supplemented with 1X Protease Inhibitor cocktail (***2.1.3.4***). Soluble extract was prepared by sonication and centrifugation as previously described for the SMARCAD1 purification (**2.7.1**). The sample was loaded through a 150mL Superloop™ onto a 5mL HisTrap™ HP column at a flow rate of 0.8mL/minute. The column was first washed in 10CV of HW-200Zn buffer (***2.1.5.11***) at a flow rate of 2mL/minute, then in HW-100Zn buffer until the conductivity trace stabilized at the baseline. The stable protein complex was eluted in a 100mL gradient of 0-100% HE-300Zn buffer (***2.1.5.4***) at a flow rate of 1mL/minute. The imidazole gradient separates the free CUE1,2 fragment from the complex, which elutes later in the gradient. The peak fractions were pooled, to which 10µL of Ulp1 (56mg/mL) was added. The sample was then dialyzed overnight at 4°C against 2L of Q-100Zn buffer (***2.1.5.8***).

Although the sample frequently became cloudy after overnight dialysis, this was not because of precipitation of the complex. Nevertheless, the sample was centrifuged at 3900×g for 30minute and further clarified through a sterile Millex®-GP 0.22µm syringe filter unit. The sample was then loaded through a 50mL Superloop™ (GE Healthcare) onto an equilibrated 8mL Mono Q® HR 10/10 column at a flow rate of 1mL/minute. The complex was eluted from the column in 1.5mL fractions using a 160mL (20CV) gradient of 0-50% Q-1000Zn buffer (***2.1.5.9***) at a flow rate of 1mL/minute. The peak fractions were pooled and concentrated to approximately 4mL using an Amicon® Ultra-15 30K MWCO spin concentrator (Millipore).

The sample was then loaded using a 10mL Superloop™ onto a 120mL HiLoad™ 16/60 Superdex® 200 gel filtration column (GE Healthcare) at a flow rate of

0.3mL/minute. The column was eluted isocratically with GF-150Zn buffer (***2.1.5.12***) over 1.1CV at a flow rate of 0.3mL/minute in 1.2mL fractions. The stable complex eluted as a single peak and was separated from the free CUE1,2 fragment. The peak fractions were pooled and concentrated to approximately 500µL (i.e. 10-16mg/mL) using an Amicon® Ultra-15 30K MWCO spin concentrator (Millipore).

2.7.6 Purification of Ulp1

The cell pellet from a 4L culture of *E. coli* expressing Ulp1 (**2.4.3**) was resuspended in 50mL of Ulp1 lysis buffer (***2.1.5.13***) supplemented with 1X Protease Inhibitor cocktail (***2.1.3.4***). Lysozyme (40mg/mL) dissolved in lysozyme reconstitution buffer was added to a final concentration of 1mg/mL. The sample became more viscous, suggesting cell lysis, following incubation at 4°C for 45 minutes. Soluble extract was prepared by sonication and centrifugation as previously described for the SMARCAD1 purification (**2.7.1**). The sample was loaded onto 5mL of pre-equilibrated Ni-NTA agarose (Qiagen) at a flow rate of 1mL/minute. The column was first washed in 8CV of Ulp1 wash buffer 1 (***2.1.5.14***), then in 8CV of Ulp1 wash buffer 2 (***2.1.5.15***) at a flow rate of 1mL/minute. Proteins were eluted from the column with Ulp1 elution buffer (***2.1.5.16***); Ulp1 eluted in a sample volume of approximately 25mL. Fractions containing Ulp1 were concentrated into an approximate total volume of 4mL using two Amicon® Ultra-15 15K MWCO spin concentrators (Millipore).

The sample was then loaded using a 10mL Superloop™ (GE Healthcare) onto an equilibrated 120mL HiLoad™ 16/60 Superdex® 200 gel filtration column (GE Healthcare) at a flow rate of 0.4mL/minute. The column was eluted isocratically with Ulp1 GF buffer (***2.1.5.17***) over 1.5CV at a flow rate of 0.25mL/minute in 1.5mL fractions. Fractions containing Ulp1 were concentrated into an approximate total volume of 1.6mL using two Amicon® Ultra-15 15K MWCO spin concentrators (Millipore). DTT and glycerol were added to final concentrations of 10mM and 10% (v/v) respectively. The protein concentration of the final sample was 56mg/mL.

2.7.7 Purification of GST-Ubiquitin Resin

The cell pellet from a 1L culture of *E. coli* expressing GST-Ub (**2.4.3**) was resuspended in 20mL of GST-L buffer (***2.1.5.18***) supplemented with 1X Protease Inhibitor cocktail (***2.1.3.4***). Following addition of lysozyme (40mg/mL) to a final concentration of 1mg/mL to the sample, it was incubated at 4°C for 45 minutes, becoming significantly more viscous. Soluble extract was prepared by sonication and centrifugation as previously described for the CUE fragment purification (**2.7.2**). The sample was mixed with 3mL of pre-equilibrated Glutathione Sepharose™ 4B (GE Healthcare) and allowed to incubate at 4°C for at least 3 hours. The column was first washed in 10CV of GST-W500 buffer (***2.1.5.19***), then in 10CV of GST-W50 buffer (***2.1.5.20***) at a flow rate of 2mL/minute. The column was finally washed in 10mL of GST-W100 buffer (***2.1.5.21***), in which it was stored at 4°C until used.

2.7.8 Preparation of Multi-Dsk Resin

Multi-Dsk resin was prepared as described in (Wilson et al., 2012). The cell pellet from a 1L culture of *E. coli* expressing GST-MultiDsk was resuspended in 40mL of STE buffer (***2.1.5.22***) supplemented with 1X Protease Inhibitor cocktail (***2.1.3.4***). Following addition of lysozyme (40mg/mL) to a final concentration of 1mg/mL to the sample, it was incubated at 4°C for 30 minutes. DTT and N-lauryl sarcosine were added to the sample to final concentrations of 5mM and 1.5% (v/v) respectively. The sample was kept on ice while being sonicated for 1 minute in 15s on/40s off cycles at 25% amplitude using a tip sonicator (Branson). The sample was spun at 16000×g for 30 minutes; the supernatant was then further filtered through a few layers of gauze. After Triton™ X-100 was added to the sample to a final concentration of 1.5% (v/v), the sample was incubated on ice for 5 minutes. The sample was mixed with 2mL of pre-equilibrated Glutathione Sepharose™ 4B (GE Healthcare) and allowed to incubate at 4°C for 4 hours. The beads were then washed in batch in 80mL of MD wash buffer 1 (***2.1.5.23***) twice. Using a peristaltic pump flowing at a rate of 2mL/minute, the beads were further washed in 100CV of MD wash buffer 1, then in 50CV of MD wash buffer 2 (***2.1.5.24***), before being equilibrated in PBS, in which it was stored until use.

2.8 Biochemical Assays

2.8.1 Reconstitution of the SMARCAD1-KAP1 Complex

180µg of purified FLAG-SMARCAD1 was combined with 180µg of purified HA-KAP1 in a 220µL binding reaction that also contained 1X SK reconstitution buffer (***2.1.6.1***) and 5mM β-ME. The binding reaction was then allowed to mix for 30 minutes at 4°C, before being added to 300µL of pre-equilibrated anti-HA (3F10) affinity matrix (Roche) and allowed to incubate at 4°C for 12 hours. The resin was washed in 500µL of 1X SK reconstitution buffer thrice, before bound proteins were eluted by incubation at 37°C for 30 minutes with 300µL of 1mg/mL HA peptide (Peptide Chemistry core facility, LRI) prepared in 1X SK reconstitution buffer. HA peptide elution was repeated again, and the two fractions were pooled. The eluate was added to 300µL of pre-equilibrated anti-FLAG® M2 affinity gel (Sigma) and incubated at 4°C overnight. The FLAG resin was washed in 500µL of 1X SK reconstitution buffer thrice, before bound proteins were eluted by incubation at 30°C for 30 minutes with 300µL of 500µg/mL FLAG peptide (Peptide Chemistry core facility, LRI) prepared in P-100 buffer (***2.1.5.6***). Although FLAG peptide elution was repeated one more time, the first fraction contained the majority of the reconstituted SMARCAD1-KAP1 complex.

2.8.2 Analytical Gel Filtration Chromatography

Analytical gel filtration chromatography was performed using a 4×300mm MAbPac SEC-1 column (Thermo Fisher Scientific) to compare the Stokes (hydrodynamic) radii of purified SMARCAD1, KAP1 and the reconstituted SMARCAD1-KAP1 complex. The different proteins samples were adjusted to the same protein concentration by addition of buffer before 45µL (375ng) of each protein was injected onto the column and eluted in P-200 GF buffer (***2.1.6.2***) over a 1.5CV isocratic gradient in 60µL fractions. A gel filtration markers kit spanning the 29-700kDa range (Sigma) was used to calibrate the column; likewise, the void volume was determined empirically using blue dextran (Sigma).

2.8.3 ATPase Assays

ATPase assays were performed as 5µL reactions containing 1X ATPase buffer (***2.1.6.3***), enzyme (typically 120nM), co-factor (typically 80nM reconstituted 30-601-30 nucleosome), and ATP (final concentration of 0.1mM). Variations to these typical values are explicitly described in relation to the specific experiment. As the purified protein samples were in P-100 buffer (***2.1.5.6***) that contains 100mM sodium chloride, in experiments where the enzyme concentration was titrated, the reactions were adjusted to the same sodium chloride concentration. ATP, as a mixture of cold ATP and ^{32}P radiolabelled ATP, was typically the final component added to the reaction. The amount of radiolabelled ATP added to each sample varied according to the specific activity of the stock – for a 5µL ATPase reaction, the amount of 6000Ci/mmol EasyTide ATP [γ-^{32}P] (PerkinElmer) added ranged from 0.05µL when used fresh, to 0.2µL after approximately 4 half-lives had elapsed. The co-factors tested in this study included single-stranded DNA (a 150-mer oligonucleotide), double-stranded DNA (annealed 150-mer oligonucleotides), and recombinant 30-601-30 human nucleosomes (generously gifted by Daniel Maskell) – recombinant human histone octamers positioned on the Widom 601 sequence and flanked by 30bp of DNA on either side.

The ATPase reactions were incubated at 37°C; typically, ATP hydrolysis was measured at two time points – 20 and 60 minutes. To stop the reaction, 2µL of the reaction was transferred to microcentrifuge tubes containing 0.5µL of 500mM EDTA. For each reaction, 2µL of the resulting mixture was spotted on PEI-impregnated cellulose TLC plates (Machery-Nagel), approximately 1cm from the bottom, and dried using a hairdryer. The plate was incubated in a TLC chamber filled with TLC running buffer (***2.1.6.4***), which travelled up the plate by capillary action, separating ATP from free inorganic phosphate in the process. The TLC plate was dried using a hairdryer before being exposed to a phosphorimager screen (GE Healthcare), which was processed into a digital image using a Typhoon™ Trio variable mode imager (GE Healthcare). ATP hydrolysis was quantified by densitometry using the ImageQuant™ TL software (GE Healthcare).

2.8.4 SMARCAD1-KAP1 Binding Assays

To interrogate the SMARCAD1-KAP1 interaction with purified full-length recombinant proteins *in vitro*, 7µg of FLAG-SMARCAD1 was mixed with 7µg of HA-KAP1 in a 280µL binding reaction containing 1X SK binding buffer (***2.1.6.5***), 0.1mg/mL BSA, and 5mM β-ME. The final concentration of sodium chloride in each binding reaction was adjusted to 200mM. The binding reactions were mixed at 4°C for 1 hour before 20µL of pre-equilibrated anti-FLAG® M2 affinity gel (Sigma) was added to each reaction. The binding reaction was immunoprecipitated with FLAG beads overnight at 4°C before being washed in 500µL of SK-200 buffer (***2.1.6.6***) three times. 30µL of 2X SDS-PAGE loading buffer (***2.1.3.1***) was added to each sample and heated at 100°C for 5 minutes. The binding assay was evaluated by SDS-PAGE and InstantBlue staining.

The binding assay involving purified fragments of SMARCAD1 and KAP1 was performed in a similar manner using the same set of buffers. However, 9.6µg of FLAG-CUE1,2 (S95-E347) was mixed with 9.6µg of KAP1 S33-K434 in a 240µL binding reaction containing 1X SK binding buffer, 0.1mg/mL BSA, and 5mM β-ME. The final concentration of sodium chloride in each binding reaction was also adjusted to 200mM. Following mixing of the binding reaction at 4°C for 1 hour, the FLAG-CUE1,2 fragment was immunoprecipitated using 20µL of pre-equilibrated anti-FLAG® M2 affinity gel (Sigma), which was incubated with the binding reaction at 4°C for 3 hours. The beads were then washed in 500µL of SK-200 buffer three times. Elution was achieved by resuspending beads in 30µL of 2X SDS-PAGE loading buffer (***2.1.3.1***) and heating the sample at 100°C for 5 minutes. The effect of ubiquitin on the SMARCAD1-KAP1 interaction was investigated by adding purified recombinant, monomeric ubiquitin (Boston Biochem) to the binding reaction. A molar ratio of ubiquitin to CUE1,2 fragment of 100:1 was readily accommodated within this assay. While it was technically feasible to test a molar ratio of 1000:1, at such high protein concentration, the sample was poorly resolved by SDS-PAGE.

To identify the region of KAP1 recognized by the first CUE domain of SMARCAD1, an affinity resin comprised of immobilized SMARCAD1 CUE1,2 fragment was prepared. 180µg of CUE1,2 fragment was added to a 300µL binding reaction

containing 1X SK binding buffer and 5mM β-ME; as before, the final salt concentration was adjusted to 200mM. As the intention was to saturate the beads with the CUE1,2 fragment, the amount of protein added was three times the binding capacity of the M2 resin (approximately 12.2nmol protein/mL resin). The beads were incubated overnight at 4°C before being washed in 500µL of SK-200 buffer twice. The beads were then equilibrated in 500µL of GST-L-Zn buffer.

Expression of GST-tagged KAP1 fragments was induced with 1mM IPTG at 30°C for 3 hours in 50mL *E. coli* cultures. Extracts were prepared by resuspending each *E. coli* pellet in 1mL of GST-L-Zn buffer (***2.1.6.7***) supplemented with 1X protease inhibitor cocktail (***2.1.3.4***) and 1mM $CaCl_2$. Following addition of lysozyme to a concentration of 2mg/mL, the samples were incubated at 4°C for 30 minutes. The viscous samples were then sonicated in the Bioruptor (Diagenode) for 6 minutes in 15s on/30s off cycles at high amplitude, before being digested at 4°C for 30 minutes with micrococcal nuclease (New England Biolabs), used at a concentration of 2000 gel units/mL. The extracts were clarified by centrifugation at 14000rpm for 30 minutes in a 5417R microcentrifuge (Eppendorf). 20µL of CUE1,2-coupled resin was used for each IP reaction, to which *E. coli* extract (2.5mg of total protein per reaction) was added and incubated at 4°C for 3 hours. The beads were washed in 500µL of SK-200 buffer thrice. Finally, the beads were resuspended in 30µL of 2X SDS-PAGE loading buffer (***2.1.3.1***) and heated at 100°C for 5 minutes to elute proteins bound to the CUE1,2-coupled resin.

2.8.5 Limited Tryptic Proteolysis

For limited tryptic digestion, a 60µL reaction containing 1X trypsin buffer (***2.1.6.8***) and 24µg of purified recombinant protein was set up. After a 10µL aliquot was collected as the “no trypsin” control, 20ng of trypsin dissolved in trypsin reconstitution buffer (***2.1.3.11***) was added to the reaction, which was incubated at 20°C. 10µL aliquots were collected after 5, 20, 60 and 150 minutes and transferred to microcentrifuge tubes containing 4X SDS-PAGE loading buffer (***2.1.3.2***) supplemented with 4X protease inhibitor cocktail (***2.1.3.4***).

Two complementary techniques were used to map the fragments arising from the limited digestion: Edman degradation and intact molecular weight mass spectrometry. For Edman degradation, digested samples were resolved by SDS-PAGE, transferred onto an Amersham® Hybond P 0.45 polyvinylidene fluoride (PVDF) membrane (GE Healthcare) that had been activated by incubation in methanol, and stained with Ponceau S (Sigma). After allowing the membranes to air-dry, the stained bands were excised. Edman degradation was performed by AltaBioscience; for each sample, 5 degradation cycles were performed.

As the buffers of the purified protein samples contain β-ME, which covalently attach to cysteine thereby introducing sample heterogeneity, the digested samples were incubated with 50mM DTT on ice for 30 minutes to remove β-ME adducts. Tryptic peptides were then removed using an Ultrafree®-CL centrifugal filter unit with a 5K MWCO (Millipore). LC/MS grade formic acid (Fisher Scientific) was then added to the sample to a concentration of at least 0.2% (v/v) for a target of approximately pH3; the pH was tested by spotting a drop of the sample on pH indicator strips. The samples were frozen and stored at -80°C until intact molecular weight mass spectrometry analysis was performed by Steve Howell (Protein Analysis and Proteomics core facility, LRI).

2.8.6 Peptide Arrays

Peptide arrays were synthesized by the Peptide Chemistry core facility (LRI) using a Multipep peptide synthesizer (Intavis AG). Peptides were spotted onto a cellulose membrane. In this study, the peptide arrays were composed of peptides 20 amino acids in length and were designed to scan the primary sequence of a protein by moving along the sequence in 1bp steps.

The dry peptide array membranes were reactivated by incubation in 50% (v/v) methanol for 10 minutes, rinsed in TBS-Tween (***2.1.3.9***) for 10 minutes, and then blocked in 4% (w/v) milk dissolved in TBS-Tween. The membranes were then rinsed in TBS-Tween for 10 minutes, before being incubated overnight at 4°C with purified protein diluted in 1% (w/v) milk; in this study, 80µg of either the wild type

CUE1,2 fragment or the version with point mutations within the first CUE domain was used. Unbound protein was washed off the membrane using TBS-Tween; three washes of 10 minutes each were performed. To determine whether the protein bound to any peptides on the array, the membrane was processed as if it were a regular Western blot. It was incubated with primary antibody diluted in 1% (w/v) milk at room temperature for 2 hours, washed for 10 minutes in TBS-Tween thrice, incubated at room temperature for 1 hour with a secondary antibody solution prepared in 1% (w/v) milk, before finally being washed in TBS-Tween three times in 10 minute cycles. The arrays were then visualized using the SuperSignal™ West Pico/Dura Chemiluminescent Substrate (Thermo Fisher Scientific) and luminescence detected with Amersham™ Hyperfilm ECL (GE Healthcare).

2.9 Crystallography Techniques

2.9.1 Crystallization Screens

The 14 commercial crystallization screens used to identify initial crystallization conditions are as follows: Natrix HT, Crystal Screen HT, Peg/Ion HT, Index HT (Hampton Research), PACT *premier*™, JCSG-*plus*™, ProPlex™, Midas™, Morpheus® (Molecular Dimensions), JBScreen Kinase, JBScreen Pentaerythritol, Cryo 1 & 2 (Jena Biosciences), Wizard 1 & 2 and Wizard 3 & 4 (Rigaku). These screens were performed in 96-well format, using the sitting drop vapour diffusion method. The Mosquito® Crystal robot (TTP Labtech) was used to mix 0.2µL of the protein solution (9-15mg/mL) with 0.2µL of the mother liquor on a platform within each enclosed well. The plates were incubated at 18°C and manually inspected for crystals using a Leica M205 C stereo microscope (Leica Microsystems) on alternate days in the first week after setting up the crystal trays, and subsequently, weekly or fortnightly for up to 2 months.

2.9.2 Optimization of Crystallization Conditions

The initial crystallization conditions identified from the screens were replicated using the hanging drop vapour diffusion method in 48-well or 24-well plates, for which 1µL of the protein solution was used per drop. The crystallization conditions

were also optimized in this format by iteratively varying each component of the mother liquor. Thus, variables that were tested included the pH, concentration of the precipitant and concentration of other ions in the mother liquor. It was also investigated whether the Additive Screen HT (Hampton Research) improved crystallization. While a 1:1 drop ratio of protein solution to mother liquor was used by default, drop ratios of 0.5:1 and 2:1 were also tested. For some conditions, microseeding was also tried.

2.9.3 Mounting of Protein Crystals

Crystals obtained in this studied were mounted by Valerie Pye. Each crystal was caught with a cryogenic loop (Hampton Research), immersed in solutions of the mother liquor with increasing concentrations of cryoprotectant, before being frozen and stored in liquid nitrogen. Of the crystallization conditions in this study, one contained PEG 400; thus, crystals that formed in that condition were frozen in mother liquor supplemented with 30% (v/v) PEG 400. For each of the two other conditions, three different cryoprotectants – 30% (v/v) PEG 400, glycerol, or ethylene glycol – were used.

Chapter 3. Results I: Purification & Mass Spectrometry Analysis of Endogenous CHD1 from Murine Embryonic Stem Cells

3.1 Aims

CHD1 belongs to the Chd1 subfamily of chromatin remodelling ATPases, which is defined by the presence of chromodomains, a conserved helicase-like ATPase domain and a DNA-binding domain that is structurally similar to the ISWI SANT and SLIDE domains (Clapier and Cairns, 2009, Ryan et al., 2011, Sharma et al., 2011). CHD1 is highly conserved, with homologues in yeast to man. Nonetheless, it has acquired metazoan-specific functions – most notably, it is essential for ES cell proliferation and for the maintenance of pluripotency, as depletion of CHD1 prevents mouse ESCs from differentiating into primitive endoderm (Gaspar-Maia et al., 2009).

While the conserved SWI2/SNF2-like ATPase domain of certain chromatin remodellers, such as that of the monomeric *Drosophila* ISWI protein, may in itself be sufficient for recognizing and remodelling nucleosomal substrates (Mueller-Planitz et al., 2013), accessory subunits frequently alter the affinity and specificity of chromatin remodellers for specific substrates (Chaban et al., 2008, Chen et al., 2013, Nguyen et al., 2013, Tosi et al., 2013). Tissue-specific subunits may, therefore, confer upon ubiquitously expressed chromatin remodelling complexes the ability to regulate the expression of specific sets of genes. Indeed, changes in composition of the mammalian SWI/SNF (BAF) chromatin remodelling complex to incorporate tissue-specific subunits allow it to function as an important developmental regulator (Lessard et al., 2007, Ho et al., 2009a, Ho et al., 2009b, Ho et al., 2011). In contrast, Chd1 has been previously characterized to be principally monomeric in budding yeast and in *Drosophila* (Tran et al., 2000, Lusser et al., 2005, Ehrensberger and Kornberg, 2011). Hence, we hypothesized that the role of CHD1 in pluripotency was acquired late in evolution and is mediated by hitherto unknown complex subunits. To address this question, we sought to purify

the native CHD1 complex from mouse ESCs and to characterize its subunits by mass spectrometry analysis.

3.2 Results

3.2.1 Conventional Protocols Fail to Extract Chromatin-Bound CHD1

To identify novel stem-cell specific subunits of the putative CHD1 complex, it was essential to purify the native, endogenous CHD1 complex from stem cells. Endogenous protein complexes have traditionally been isolated from native tissues by purification over an extensive series of columns while tracking a specific biochemical activity. However, this approach was judged to be impractical in this situation where the amount of source material – mouse ESCs – is limited. Instead, we decided to purify CHD1 from a mouse ESC line that expresses epitope-tagged CHD1. The comparatively high rate of homologous recombination in ESCs was exploited to recombine a FLAG-TEV-HA epitope tag into the 3' end of the *CHD1* gene (Figure 3.1A), such that CHD1 with a C-terminal epitope tag was expressed from its native promoter at endogenous level. This minimises the probability of non-specific interactions arising from an excess of overexpressed CHD and as such, it is highly likely the composition of the tagged CHD1 complex will retain its native stoichiometry. As these epitope tags are relatively small, it is unlikely that they interfere with protein-protein interactions involving CHD1; however, this possibility was not formally tested here.

To assess the feasibility of purifying CHD1 from mouse ESCs, the relative expression level of CHD1 in mouse ESCs was compared to that of other CHD proteins. By a considerable margin, CHD4, the catalytic subunit of the NuRD complex, was the most abundant member of this protein family in mouse ESCs. After CHD4, however, CHD1 was probably the next most abundant of the CHD proteins. As expected, subcellular fractionation indicated that CHD1 localized primarily to the nucleus, as there was significant enrichment of CHD1 in the nuclear fraction (Figure 3.1B). On the basis of this data, we felt that the purification of CHD1 from mouse ESCs would be possible, albeit with limited yields.

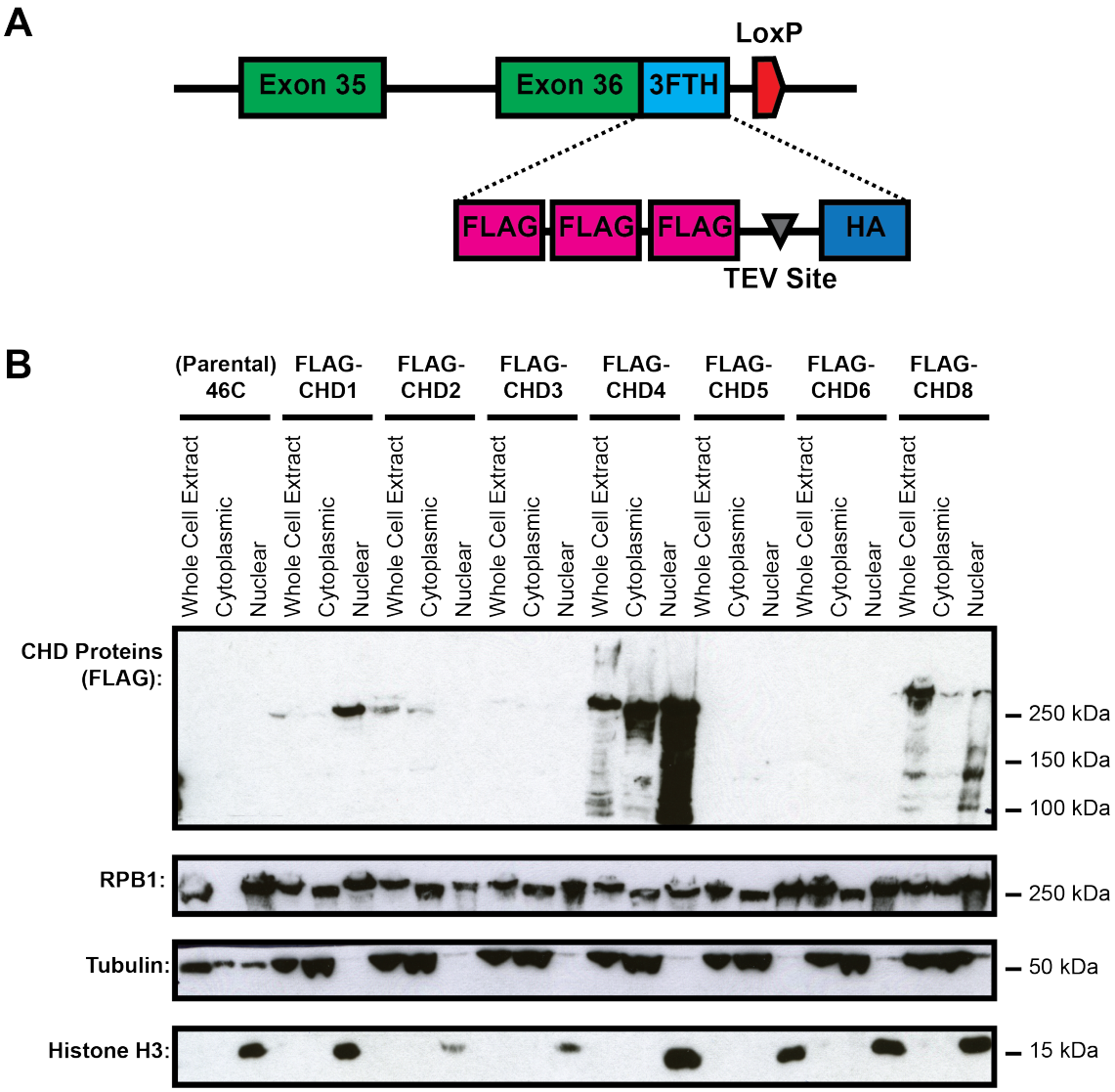


Figure 3.1 – Comparison of the Expression Levels of the CHD Proteins in mESCs

A. This schematic illustrates the modified *CHD1* gene locus following homologous recombination. A 3FTH tag – comprised of three FLAG tags, an intervening TEV cleavage site and an HA tag – is introduced into the end of exon 36, the last exon of the *CHD1* gene.

B. These mouse ESCs express C-terminally tagged CHD proteins at endogenous levels. The expression levels of the different CHD proteins relative to each other were compared by a Western blot against the FLAG tag. After CHD4, CHD1 is one of the more abundant members of this protein family in mouse ESCs. Subcellular fractionation indicates that CHD1 is enriched in the nuclear fraction.

To purify CHD1, we decided to apply a protocol previously optimized in our lab for the purification of novel interactors of RNA polymerase II (Aygun et al., 2008). Arguably the most important aspect of this protocol is that subcellular fractionation is performed prior to affinity purification, not only to reduce the sample complexity and to enrich for the protein-of-interest, but also to focus on interactions that occur

on chromatin, which have the highest likelihood of being of functional relevance. In brief, the subcellular fractionation protocol involves the isolation of nuclei by hypotonic lysis of mammalian cells, subsequently followed by preparation of a chromatin fraction by homogenization, sonication and centrifugation of the nuclei. Chromatin-bound proteins are then extracted from the chromatin pellet by digestion with benzonase, a promiscuous endonuclease that digests both RNA and DNA, into a buffer containing 150mM NaCl – roughly physiological ionic strength – to preserve as many protein-protein interactions as possible.

CHD1 was recovered almost entirely in the chromatin pellet (Figure 3.2A). However, digestion of the chromatin pellet with typical concentrations of benzonase failed to render CHD1 soluble (Figure 3.2A). The inability of benzonase digestion to extract CHD1 from the chromatin pellet probably does not reflect insufficient enzyme being added to the sample since high concentrations of benzonase and micrococcal nuclease, a nuclease that preferentially cleaves the linker DNA between nucleosomes, were similarly ineffectual (Figure 3.2B). Benzonase was confirmed to be active under these conditions – agarose gel electrophoresis indicated that most of the DNA recovered from digested chromatin samples by phenol-chloroform extraction was in the form of fragments less than 500 base pairs in length. It is known that it can be particularly challenging to extract nuclear matrix components and chromatin-associated proteins, especially those associated with highly compacted heterochromatin (Torrente et al., 2011). In that regard, the particularly tight association with chromatin demonstrated by CHD1 was rather surprising since it has been reported to associated specifically with euchromatic promoters – regions of the genome that are less compacted (Gaspar-Maia et al., 2009). Due to the intractability of extracting CHD1 from chromatin by conventional methods, alterative approaches had to be explored.

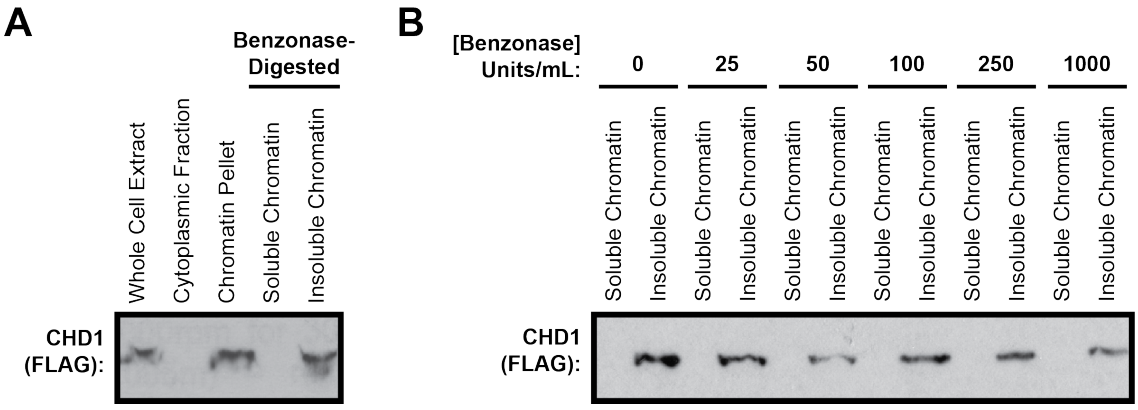


Figure 3.2 – Benzonase Digestion Fails to Extract CHD1 from Chromatin
A. Following subcellular fractionation, CHD1 is recovered almost entirely in the chromatin pellet. However, digestion of the chromatin sample with a typical concentration of benzonase, a promiscuous endonuclease, fails to release CHD1 into the soluble fraction.
B. CHD1 is highly resistant to being extracted from chromatin by benzonase digestion, as even very high concentrations of nuclease are ineffective.

3.2.2 Extraction of Chromatin-Bound Proteins by Heparin

Heparin is a polyanionic polysaccharide; in the sense that it possesses a high density of negative charge, it is similar to DNA. Indeed, in some scenarios, for example as a chromatographic substrate, it can even be regarded as a DNA-mimetic. In addition to its medical application as an anticoagulant, it has been previously used for biochemical purposes, including for the purification of nuclear membranes with minimal nucleic acid and protein contamination (Bornens, 1973, Bornens, 1977, Courvalin et al., 1982). In this context, heparin preferentially solubilizes non-histone, chromatin-bound proteins, with solubilization of chromatin occurring at higher heparin concentrations. Histones can even be induced to dissociate from chromatin if the quantity of heparin in the sample is greater than the amount of DNA (Courvalin et al., 1982). As DNA binds histones with high affinity, we reasoned that if histones could be solubilized in the presence of heparin, then it should similarly be possible to extract CHD1 from chromatin.

Addition of heparin to the resuspended chromatin pellet perceptibly altered the macroscopic appearance of the sample, which became more clarified. As reported, solubilization of chromatin-bound proteins occurred upon addition of heparin (Figure 3.3).

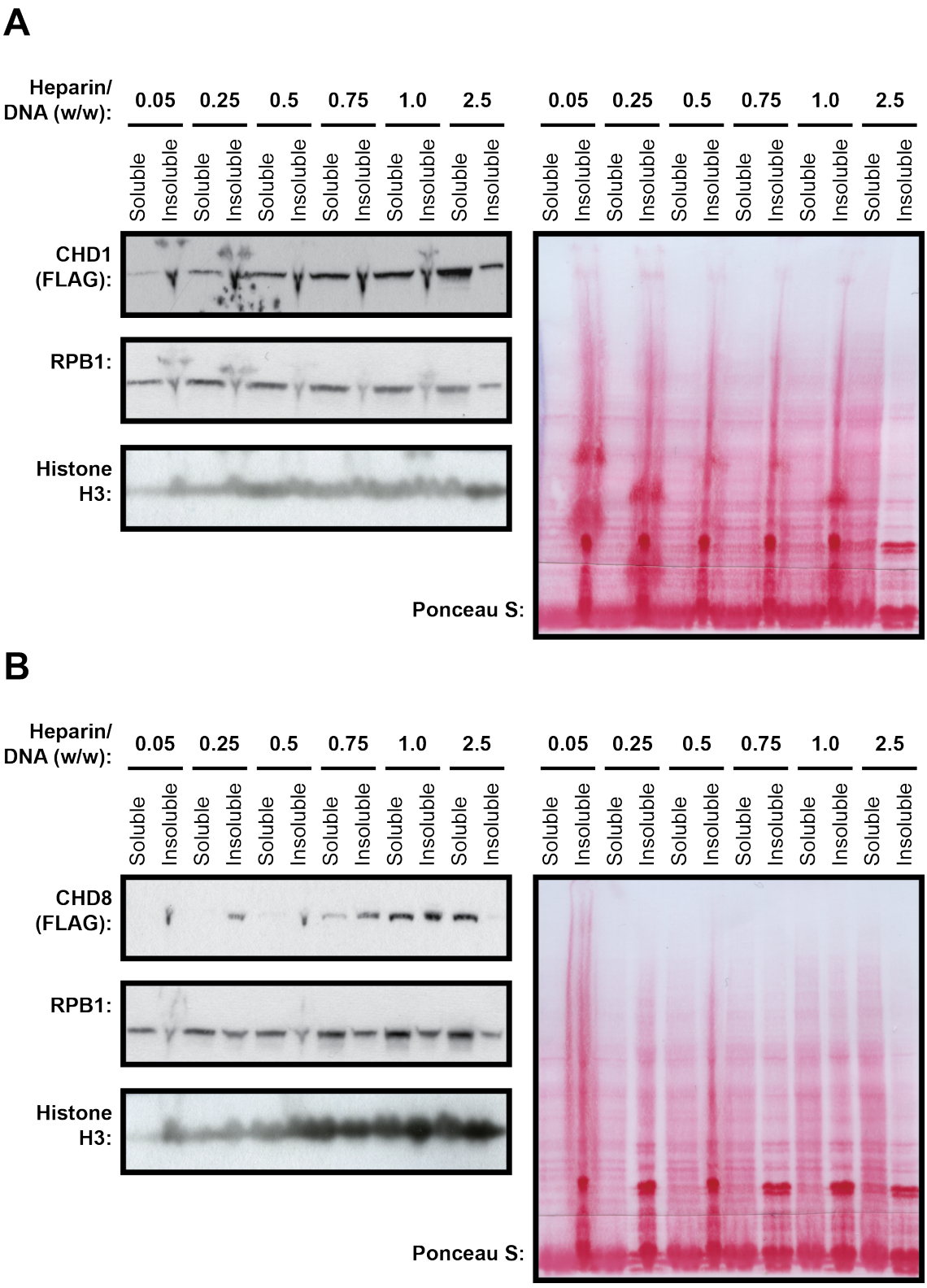


Figure 3.3 – Extraction of CHD1 & CHD8 from Chromatin by Heparin

A. Chromatin-bound proteins, including CHD1 and RPB1, are solubilized by heparin in a concentration-dependent manner (Ponceau S). At the highest heparin concentration tested, relatively few proteins apart from histones persist in the insoluble chromatin pellet. Interestingly, optimal extraction of RPB1 is achieved at a lower heparin concentration than that required for CHD1.

B. CHD8, another CHD protein that appears to associate tightly with chromatin, can also b extracted from the chromatin pellet by heparin in a concentration dependent manner.

The amount of heparin required to extract a specific protein from chromatin varied from protein to protein. For example, extraction of CHD1 or CHD8 from chromatin requires higher heparin concentrations than RPB1, the largest subunit of the RNAPII complex. At a heparin to DNA ratio (by weight) of 2.5, the majority of CHD1 and CHD8 are soluble (Figures 3.3A & B respectively). In contrast, a heparin to DNA ratio (w/w) of 0.75 sufficed for optimal extraction of RPB1, further increases in heparin concentration did not significantly increase the amount of RPB1 extracted from chromatin (Figure 3.3). Interestingly, the selective ability of the initiating, but not the elongating, form of RNAPII to bind to a heparin column has previously been noted (Gnatt et al., 1997). Thus, it is likely that the residual amount of chromatin-bound RPB1 following heparin extraction represents RPB1 molecules incorporated into elongating RNAPII complexes.

While heparin extracts proteins from chromatin in a concentration dependent manner, the key determinant is not the absolute heparin concentration of the sample. Instead, the amount of heparin relative to the amount of DNA in the sample determines the extent to which chromatin-bound proteins are solubilized (Figure 3.4A). Moreover, it was observed that the addition of herring sperm DNA to the resuspended chromatin pellet was sufficient to solubilize some chromatin-bound proteins, including CHD1 (Figure 3.4B). Taken collectively, these data suggest that heparin exploits its chemical similarities to DNA to directly compete proteins off chromatin. Importantly, the relatively indiscriminate manner in which proteins are extracted from chromatin by heparin suggests that this technique is broadly applicable for the preparation of extracts of chromatin-associated proteins.

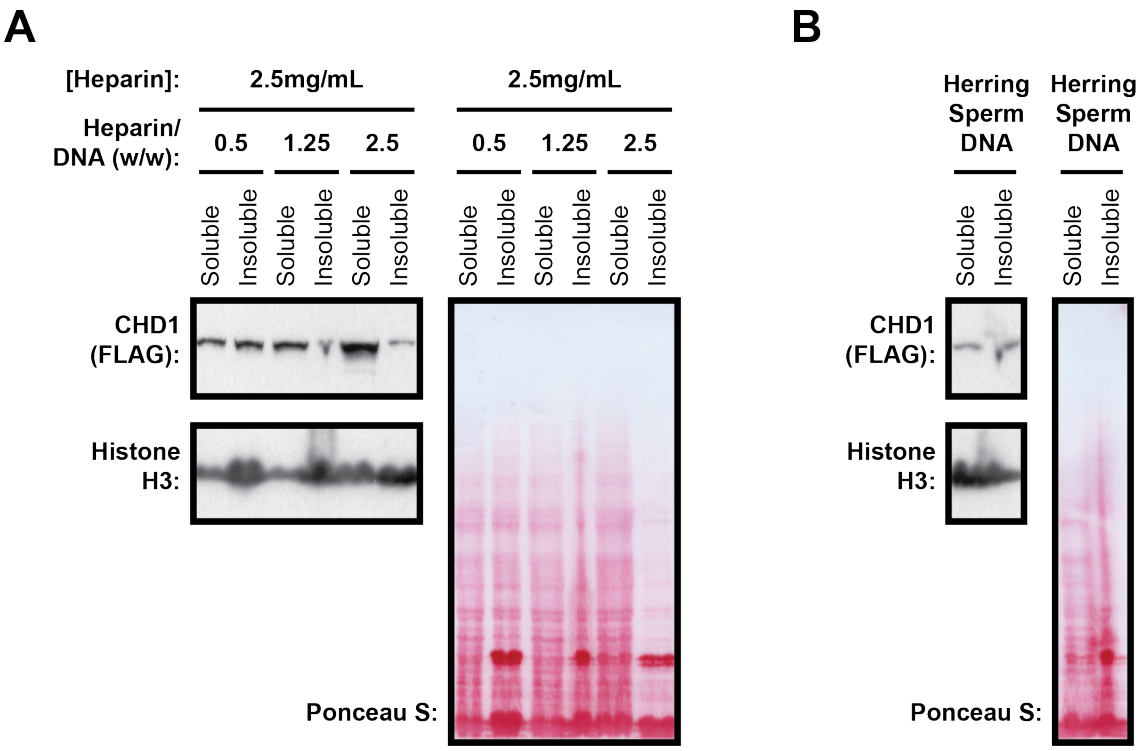


Figure 3.4 – Characterization of Heparin Extraction of Chromatin-Bound Proteins

A. It is the ratio of the amount of heparin to DNA in the sample, rather than the absolute heparin concentration, that determines the extent of solubilisation of chromatin-bound proteins. Each condition tested contains equal amounts of a chromatin sample, and thus comparable starting amounts of DNA. As these samples were diluted with different amounts of buffer, when heparin was added to a final concentration of 2.5mg/mL, each condition had a different ratio of heparin to DNA. As the heparin:DNA ratio increases, the amount of total protein (Ponceau S) and CHD1 extracted increases.

B. Herring sperm DNA is similarly capable of solubilizing chromatin-bound proteins and histones. Hence, it is probable that heparin relies upon its similarity to DNA to extract proteins from chromatin, possibly by direct competition.

3.2.3 Purification of RNAPII & CHD1 from Mouse ESCs

While the heparin extraction technique should be broadly applicable for the purification of most chromatin-bound proteins, it may be especially valuable when attempting to purify such proteins under low stringency conditions since heparin extraction remains efficient in buffers of low ionic strength designed to preserve weak or transient protein interactions. To confirm that heparin extraction is compatible with subsequent protein purification, we decided to purify the RNA polymerase II core complex and interactome from mouse ESC chromatin extract prepared with heparin.

RNAPII is essential in cells as it is responsible for the transcription of protein-coding genes. The core RNAPII complex consists of 12 subunits and displays high evolutionary conservation. Nonetheless, Gdown1 has only relatively recently been identified to be a novel subunit of a proportion of mammalian RNAPII complexes (Hu et al., 2006). Thus, we also wanted to test the, admittedly unlikely, hypothesis that the RNAPII complex may possess stem cell-specific subunits.

RNAPII was purified from mouse ESC chromatin extract by immunoaffinity chromatography using the 8WG16 monoclonal antibody, which was raised against the unphosphorylated RPB1 C-terminal domain (CTD) heptapeptide. Importantly, the 8WG16 antibody is a polyol-responsive antibody; thus, bound RNAPII can be induced to dissociate from the antibody column by a high polyol concentration (Thompson et al., 1990). Mammalian RNAPII has been noted to elute poorly from the 8WG16 antibody column; presumably, it binds to the column with high avidity due to the 52 heptapeptide repeats present in the CTD of mammalian RPB1. Hence, several elution conditions were tested, including elution with a CTD peptide and various buffers containing propylene glycol, a polyol. The best elution was observed with an excess (i.e. 1mg/mL) of CTD peptide, though 40% (v/v) propylene glycol in the presence of 500mM ammonium sulphate also succeeded in eluting some RNAPII (Figure 3.5A).

Western blot analysis confirmed the presence of RPB1 and RPB2, two integral subunits of the RNAPII complex, in the eluates from the 8WG16 column, suggesting that RNAPII was successfully purified from mouse ESC chromatin extract (Figure 3.5B). Silver staining of the eluates revealed more than twelve bands, indiating that other interactors beyond the core RNAPII complex were recovered (Figure 3.5C). However, tandem liquid chromatography-mass spectrometry (LC-MS) analysis of the high stringency RNAPII purification failed to identify any novel stem cell-specific subunits.

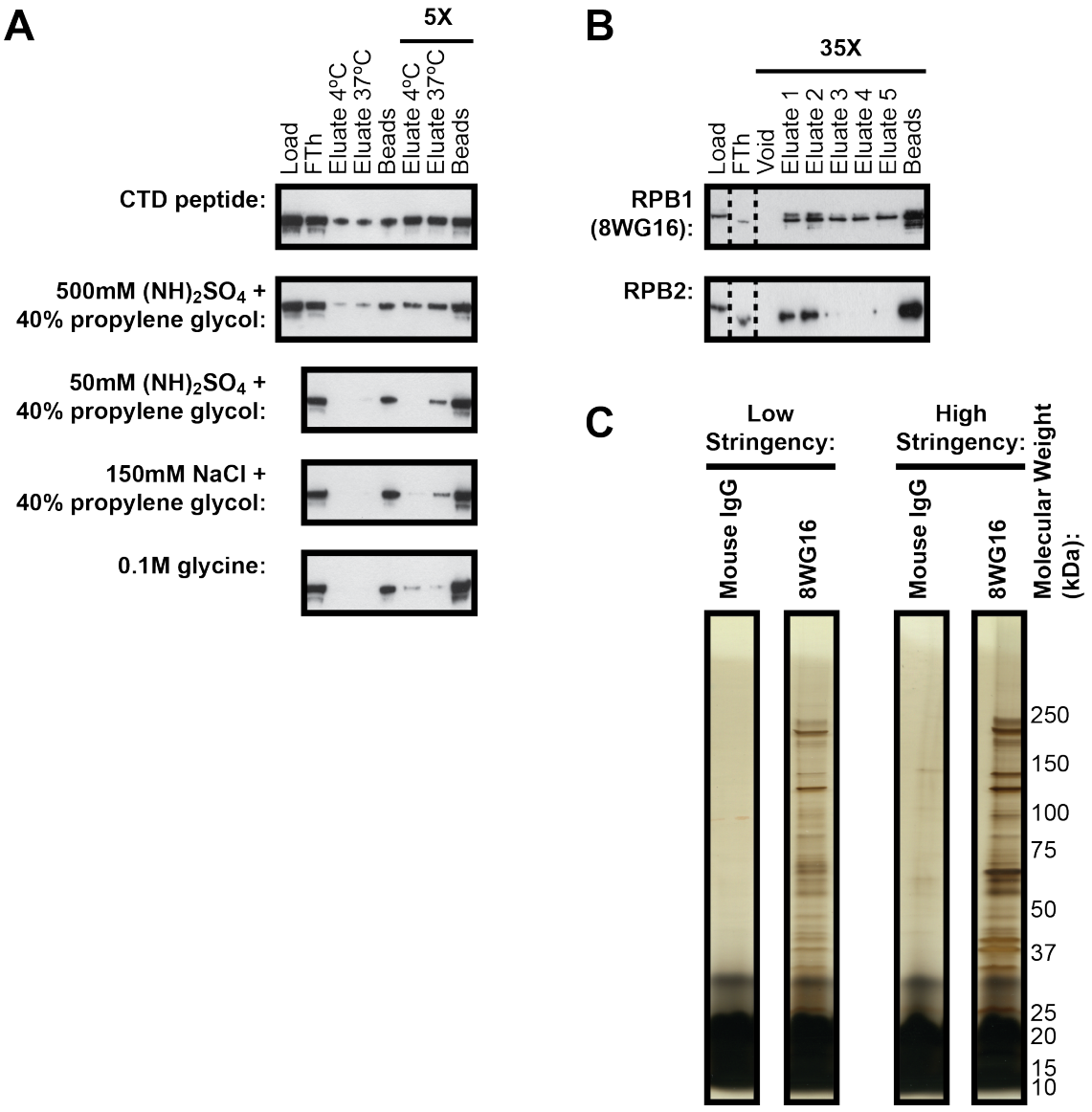


Figure 3.5 – Purification of the RNAPII Interactome from Mouse ESCs

A. Various strategies to elute the RNAPII from the 8WG16 antibody column were tested in a small-scale experiment. This Western blot against RPB1 (4H8) shows that an excess (1mg/mL) of CTD peptide is capable of competitively eluting a significant amount of RPB1 from the column. In the presence of 500mM ammonium sulphate, 40% (v/v) propylene glycol is also capable of eluting some RPB1 from the column.

B. RNAPII was purified by 8WG16 affinity chromatography from mouse ESC chromatin extract, prepared using the heparin solubilization protocol described above. RPB1 and RPB2 were efficiently depleted from extract by the antibody column. The RNAPII complex was eluted from the column by a combination of CTD peptide and 40% (v/v) propylene glycol. High and low stringency purifications were performed in parallel to isolate the RNAPII core complex and interactome respectively. While the analysis displayed is derived from the high stringency purification, it is representative of both. The dotted line denotes cropping of the image to omit irrelevant lanes.

C. The staining pattern of the eluates from the 8WG16 antibody column indicates that additional proteins were recovered in addition to the RNAPII complex. The intense smear at the bottom of the gel represents the CTD peptide used for elution. Less RNAPII is recovered from the low stringency purification than from the high stringency one due to reduced elution efficiency in a buffer of lower ionic strength.

A low stringency purification of RNAPII designed to preserve weaker interactions by omitting washes of high ionic strength was performed in parallel to define its interactome in mouse ESCs. The yields from the low stringency purification were, however, considerably less than that for the high stringency purification (Figure 3.5C). This is due to the elution from the 8WG16 affinity column being less efficient for the low stringency purification than for its high stringency counterpart. Indeed, it was previously observed that elution by propylene glycol in the context of a low ionic strength buffer was significantly less effective than if a buffer of high ionic strength were used (Figure 3.5B). Nonetheless, the low stringency RNAPII purification was analysed by LC-MS, revealing relative conservation of the RNAPII interactome between ES cells and other differentiated cells. For example, well-characterized interactors of RNAPII such as the FACT complex, Spt5 and RECQL5 were detected.The co-purification of these interactors with RNAPII suggests that heparin does not disrupt protein- protein interactions in spite of its highly polyanionic nature. This data, therefore, validates the use of heparin extraction in conjunction with low stringency affinity purifications to define the interactome of specific chromatin-associated proteins. Thus, we proceeded to purify CHD1 under both high and low stringency conditions to identify any complex subunits, and to characterize its interactome respectively (Figure 3.6).

For the high stringency purification, chromatin extract from mouse ESCs expressing epitope-tagged CHD1 was loaded sequentially onto FLAG and HA affinity columns (Figure 3.6A). A control ‘mock’ purification was performed in parallel using chromatin extract prepared from parental 46C mouse ESCs. Extensive high (750mM sodium chloride) and low (50mM sodium chloride) washes were performed to disrupt ionic and hydrophobic interactions respectively. The samples eluted off the HA columns by cleavage with the TEV protease were resolved resolved by SDS-PAGE and visualized by silver staining.

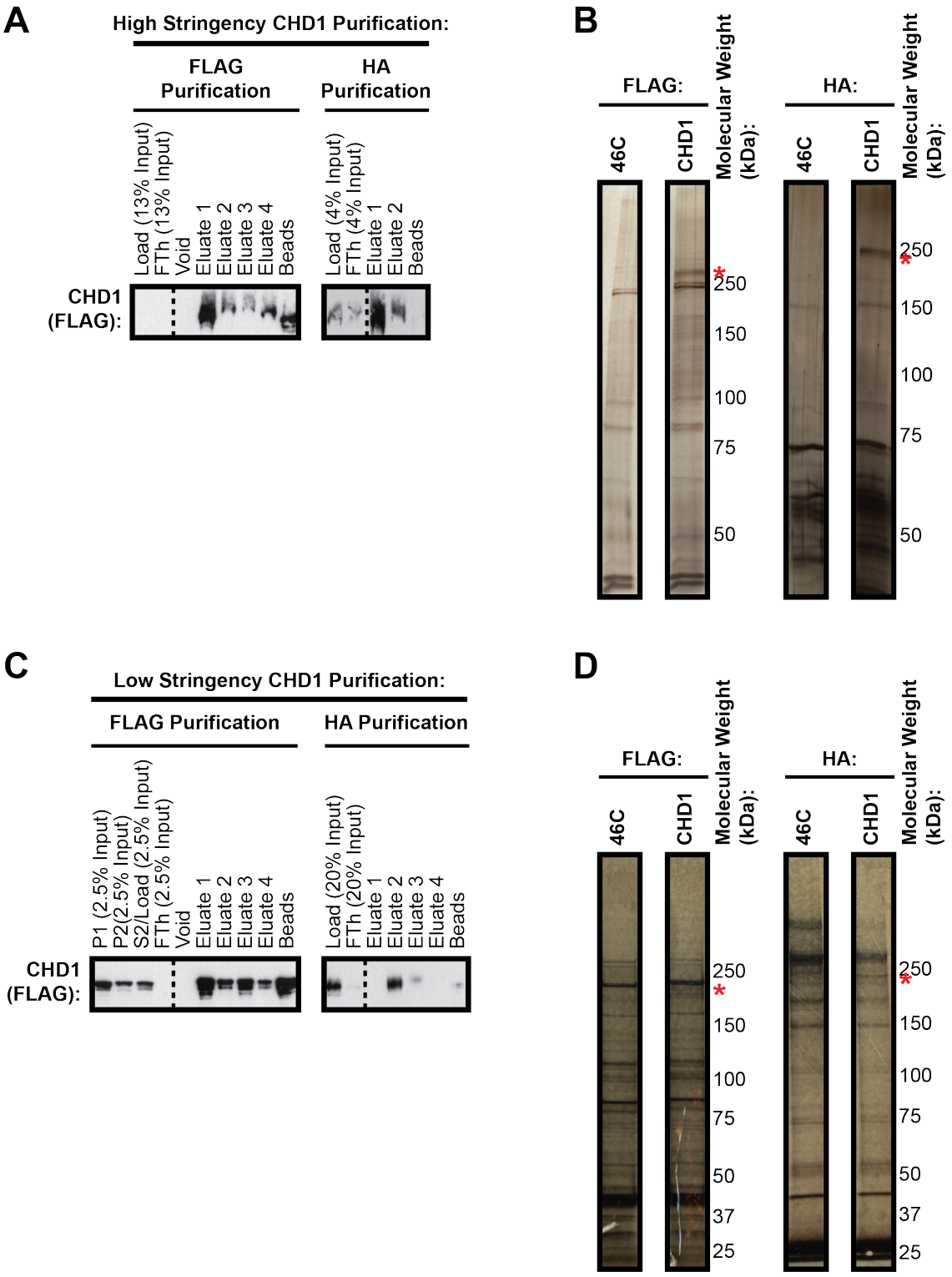


Figure 3.6 – High & Low Stringency Purifications of CHD1 from Mouse ESCs

A. The endogenous CHD1 core complex was purified from mouse ESC chromatin extract by sequential FLAG and HA affinity chromatography. For the high stringency purification, washes of up to 750mM NaCl were used, such that only strong interactions would be retained. As a control, a parallel purification was performed using chromatin from parental 46C mouse ESCs, which did not express any tagged CHD protein.

B. The eluates from the high stringency purification are of a high purity, as assessed by silver staining. Similar to its budding yeast and *Drosophila* orthologues, mouse CHD1 does not form a multi-subunit protein complex. The red asterisks indicate bands that probably correspond to CHD1.
C. To define the interactome of CHD1, CHD1 was similarly purified as before, though the entire purification was performed under mild conditions to retain even weak interactions.
D. Based on the control 46C purifications, it is clear that purification under mild conditions results in significant background contamination. It is impossible to clearly identify CHD1-specific interactors by silver staining, though as before, the red asterisks indicate bands that probably correspond to CHD1.

A unique band migrating near the 250kDa molecular weight marker is present only in the CHD1 but not the control sample almost certainly corresponds to CHD1 (Figure 3.6B). However, there were no other bands of similar intensity specific to the tagged sample. Correspondingly, while LC-MS analysis readily identified CHD1 in only the tagged sample, with over 350 assigned spectra, no other proteins in the same sample were detected with comparable frequency. Thus, CHD1 probably does not exist as part of a stable stoichiometric, multi-subunit protein complex in mouse ESCs. Indeed, the CHD1 orthologues in budding yeast and *Drosophila* are also monomeric, suggesting that the additional functions of mammalian CHD1 are unlikely to be because of the acquisition of novel complex subunits (Tran et al., 2000, Lusser et al., 2005, Ehrensberger and Kornberg, 2011).

We then proceeded to perform a low stringency purification of CHD1 to define its interactome, including lower affinity or transient protein interactors that may nevertheless be important regulators of CHD1. The tandem affinity purification was adapted to feature mild washes of physiological ionic strength (150mM sodium chloride) to preserve more protein-protein interactions (Figure 3.6C). The numerous bands observed by silver staining in the control 46C purification reveal that significant non-specific background contamination was associated with the low stringency purifications (Figure 3.6D). Consequently, it is difficult to confidently nominate a specific band to be CHD1, much less identify specific interaction partners. LC-MS analysis identified hundreds of proteins in both samples. The relative abundance of a protein was assessed by comparing the intensity-based label-free absolute-quantification (iBAQ) scores associated with a given protein for the tagged-CHD1 sample to that of the control sample. A list of proteins enriched in the CHD1 sample over background levels was compiled (Table 3.1). These CHD1-enriched proteins were subjected to unbiased gene ontology analysis; amongst the

top ten most highly enriched terms, four terms relate to transcription or chromatin remodelling, suggesting that the proteins identified to be enriched in the CHD1 sample over background may perhaps be of biological relevance (Table 3.2). Nevertheless, these data were interpreted with caution since CHD1 could not be efficiently purified from ESCs – the silver-stained gels lacked a clear and dominant CHD1 band (Figure 3.6D). Unsurprisingly, a low signal-to-noise ratio plagued the results of the mass spectrometry analysis (Table 3.1).

Name	Enrichment	iBAQ Score		Spectral Count		PEP Value
		CHD1	Control	CHD1	Control	
Gtf2i	2.98	20.3	6.8	10	0	9.61E-23
Chd1	2.66	23.3	8.8	367	1	0
COX2	2.62	17.8	6.8	4	0	1.02E-04
Rab1;Rab1A	2.58	17.5	6.8	7	0	8.75E-15
Ndufa13	2.52	17.1	6.8	2	0	5.74E-06
L1Md-A1	2.51	17.0	6.8	4	1	4.80E-04
Gnb2	2.47	16.8	6.8	3	0	1.31E-08
Rab6a;Rab6b;Rab39;Rab39a	2.46	16.7	6.8	4	0	6.17E-11
Atp5h	2.45	16.7	6.8	3	0	2.50E-15
2700060E02Rik	2.45	16.6	6.8	4	0	1.85E-27
Dimt1	2.45	16.6	6.8	5	0	5.66E-12
Hat1	2.45	16.6	6.8	5	0	2.43E-31
Sf3b14	2.44	16.6	6.8	2	0	7.29E-15
Hsd17b10	2.44	16.6	6.8	4	0	6.23E-12
Suclg1	2.42	16.4	6.8	3	0	1.75E-25
Acat1	2.39	16.3	6.8	4	0	1.39E-09
Cdh1	2.38	16.2	6.8	7	0	9.14E-18
Rfc5	2.37	16.1	6.8	3	0	1.64E-08
Fam98b	2.36	16.1	6.8	2	0	1.64E-10
Cope	2.36	12.2	6.8	1	0	4.65E-03
Rrp7a	2.36	16.0	6.8	2	0	1.51E-03
Nup43	2.35	16.0	6.8	2	0	2.41E-06
Timm13	2.35	16.0	6.8	1	0	3.67E-09
Gnb1	2.35	15.9	6.8	1	0	4.07E-09
Dnaja2	2.34	15.9	6.8	4	0	2.47E-11
Acadm	2.34	15.9	6.8	4	0	8.33E-15
Srrt	2.33	15.8	6.8	18	0	8.99E-39
Lig1	2.32	15.8	6.8	12	0	5.15E-27
Kif22	2.32	15.8	6.8	8	0	2.14E-20
Aimp2	2.32	15.8	6.8	2	0	1.11E-11
Pdk1	2.32	15.8	6.8	3	0	2.47E-06
Exosc2	2.32	15.8	6.8	3	0	6.07E-08
Sdhb	2.31	15.7	6.8	5	0	2.26E-08
Txn1;Txn	2.31	15.7	6.8	1	0	2.65E-03
Dnajb11	2.31	15.7	6.8	3	0	3.51E-22
Ppp1ca	2.30	15.6	6.8	1	0	1.45E-07
Dstn	2.30	15.6	6.8	2	0	9.32E-04
Mrps30	2.30	15.6	6.8	5	0	9.90E-06
Nup93	2.29	15.6	6.8	8	0	9.74E-20

Smarcd1	2.29	15.6	6.8	4	0	8.33E-32
Bcas2	2.29	15.6	6.8	1	0	7.37E-07
Armc10	2.29	15.5	6.8	2	0	1.50E-05
Got2	2.28	15.5	6.8	5	0	1.50E-12
Tcea3	2.28	15.5	6.8	2	0	1.95E-11
Ass1	2.28	15.5	6.8	2	0	9.72E-06
Hadhb	2.28	15.5	6.8	5	0	1.65E-09
Uba1	2.28	15.5	6.8	12	0	7.39E-30
Actr1a;Actr1b	2.27	15.5	6.8	1	0	1.77E-03
Dpm1	2.27	15.4	6.8	2	0	6.77E-04
Plk1	2.26	15.4	6.8	4	0	4.59E-26
Trim24	1.50	14.1	9.4	7	1	1.63E-07

Table 3.1 – The Most Highly-Enriched CHD1 Interactors
The level of enrichment was calculated by normalizing the iBAQ score of a given protein for the tagged-CHD1 sample to that for the control sample. The fifty most highly enriched proteins are listed in this table. An iBAQ score of 6.8 (the score of the least abundant protein detected in this sample) was arbitrarily assigned to all proteins without detectable spectra. The iBAQ score takes into account the size of the protein; hence, they do not directly correlate with the spectral count. The lower the PEP value, the higher the likelihood of accurate identification of the protein.

	Gene Ontology Category	Enrichment Factor	P value	Benjamini Hochberg FDR
1.	Holo TFIIH complex	220	1.04E-05	0.019302
2.	Photoreceptor outer segment	220	1.04E-05	0.019349
3.	mRNA cap binding complex	220	1.04E-05	0.01949
4.	Lamellipodium membrane	220	1.04E-05	0.019585
5.	Early endosome membrane	220	1.04E-05	0.019633
6.	Integrin complex	220	1.04E-05	0.019828
7.	Mitochondrial proton-transporting ATP synthase complex	220	1.04E-05	0.019976
8.	Cdc73/Paf1 (transcription elongation factor) complex	146.67	7.09E-08	0.00049429
9.	Heterotrimeric G-protein complex	110	2.84E-07	0.0014438
10.	SWI/SNF-type complex	110	2.84E-07	0.0014829

Table 3.2 – Gene Ontology Enrichment Analysis of the CHD1 Interactome
Gene ontology analysis was performed on the proteins enriched in the CHD1 sample over background levels using the Fisher exact test. The duplicated terms were either omitted or merged. The ten most highly enriched terms are listed in this table; of these four terms (highlighted) relate to transcription or chromatin remodelling suggesting that this data set is probably of biological relevance.

Although 367 spectra were assigned to CHD1, most of the other proteins were identified from a few peptides (Table 3.2). Indeed, none of the hits was specifically identified in such a large quantity that they could be incontrovertibly be regarded as a genuine CHD1 interactor. Thus, additional selection criteria were introduced – for example, the list of CHD1-enriched hits was cross-referenced to lists of common mass spectrometry contaminants. Based on their known biological functions, the

hits were also subjectively evaluated for a plausible functional connection to CHD1. Unfortunately, none of the most highly enriched hits convincingly fulfilled all of these criteria.

3.2.4 Validation of TIF1α (TRIM24) as an Interactor of CHD1

Although TIF1α was not highly enriched (1.5 fold enrichment) in the CHD1 sample over background, it was the most interesting of the putative CHD1 interactors. TIF1α is a transcriptional co-regulator, encoded by the TRIM24 gene, that preferentially binds to histone H3 tails that are unmethylated at K4 and acetylated at residue K23 (H3K4me0K23ac) via its PHD-bromodomain (Tsai et al., 2010). Additionally, it is possible that its RING domain confers upon it E3 ubiquitin ligase activity. Incidentally, amongst the multiple nuclear receptors regulated by TIF1α is the retinoic acid receptor, RXRα, whose signalling can induce differentiation of ESCs (Fraser et al., 1998).

To validate the possible CHD1-TIF1α interaction, it was investigated whether CHD1 and TIF1α co-immunoprecipitate. Indeed, TIF1α was enriched over background levels when CHD1 was pulled down (Figure 3.7A). Conversely, some CHD1 was co-immunoprecipitated by TIF1α. As this interaction appears relatively weak, additional controls were performed to confirm its specificity – for example, the possibility of TIF1α binding non-specifically to the FLAG antibody was excluded (Figure 3.7B). Nevertheless, while this interaction is probably specific, it must be noted that it only involves a small proportion of the total population of CHD1 or TIF1α. Thus, the CHD1-TIF1α interaction represents a relatively minor subpopulation of the total population of either protein, possibly because the interaction is weak or occurs only transiently.

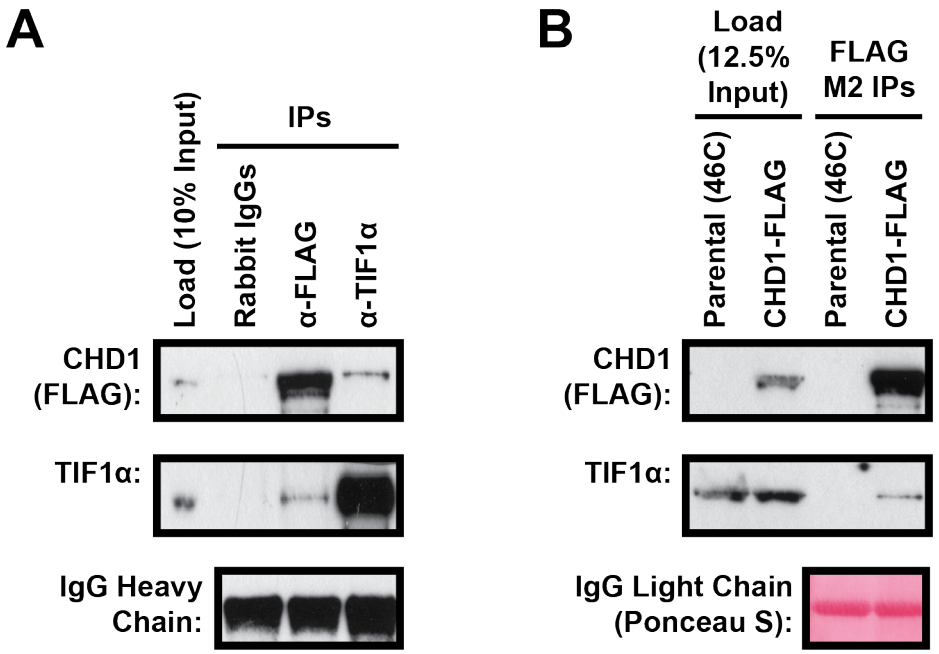


Figure 3.7 – CHD1 & TIF1α co-Immunoprecipitate

A. Immunoprecipitation of epitope-tagged CHD1 by FLAG IP with Protein G Dynabeads® (Invitrogen) resulted in co-immunoprecipitation of TIF1α. The reciprocal result was also observed – TIF1α pull-down resulted in enrichment of CHD1 compared to the IgG control. While the co-immunoprecipitation is detectable, only a small proportion of the total amount of CHD1 or TIF1α is ever pulled down by its counterpart, indicating that the CHD1-TIF1α interaction represents a minority subpopulation.

B. IP of tagged CHD1 with FLAG M2 agarose also achieved enrichement of TIF1α. However, co-IP of TIF1α does not occur when parental 46C ESC extract was used as the load, indicating that TIF1α does bind non-specifically to the FLAG antibody.

CHD1 has been implicated in the maintenance of pluripotency. Indeed, CHD1 is preferentially expressed in mouse ESCs, with downregulation of CHD1 protein levels occurring upon differentiation of the ESCs into embryoid bodies, three-dimensional cellular aggregates consisting of precursor cells of various lineages (Figure 3.8). If TIF1α is mechanistically linked to CHD1's role in pluripotency, it might be expected that the expression of TIF1α is also influenced by pluripotency. In fact, TIF1α expression is rapidly and significantly downregulated upon induction of differentiation of ESCs, achieved by transferring ESCs into suspension culture in the absence of leukaemia inhibitor factor (LIF), a cytokine that inhibits differentiation (Figure 3.8B). Thus, while certainly not definitive evidence, this result suggests that TIF1α might have a functional role in the pluripotent cell state.

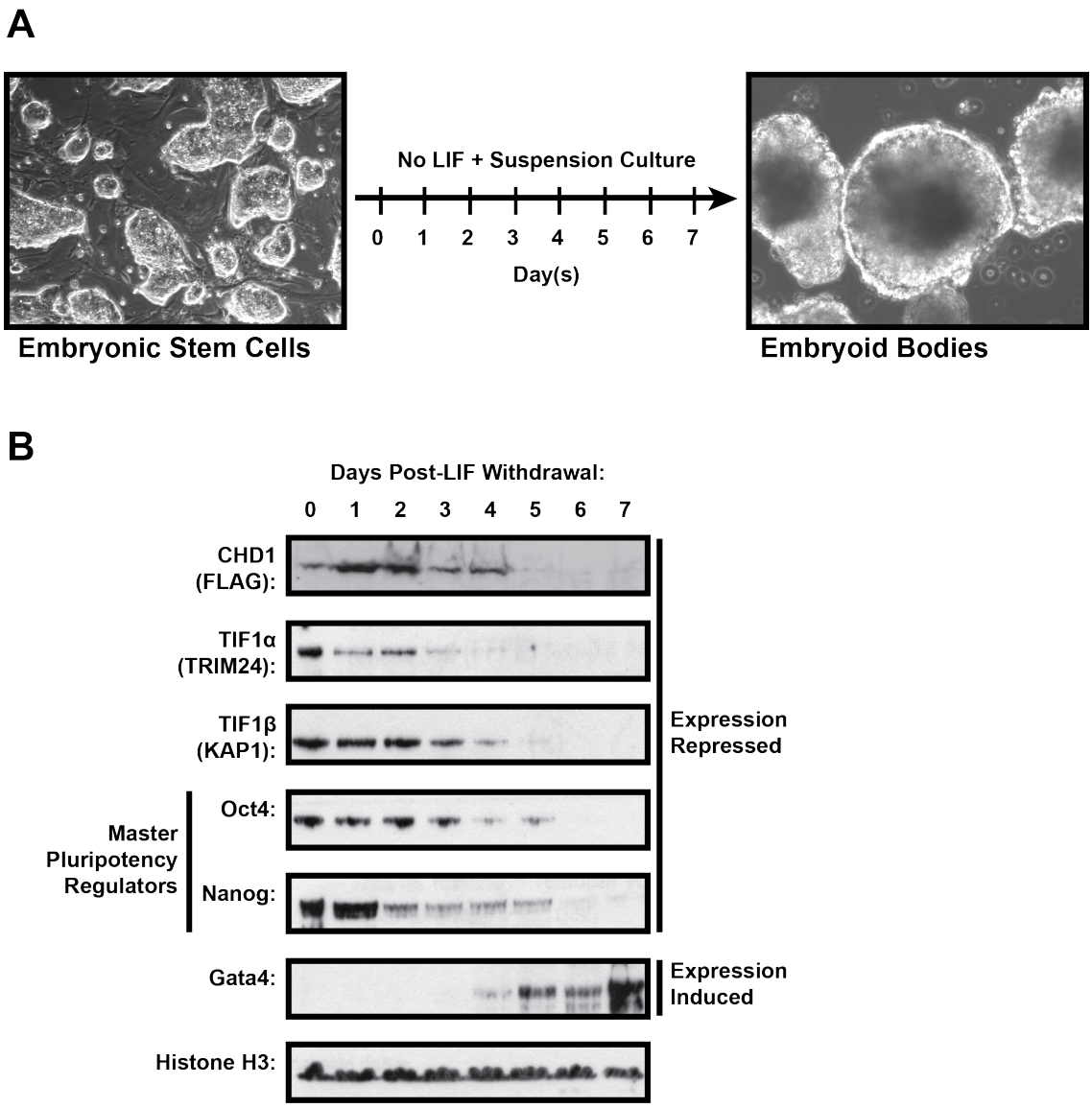


Figure 3.8 – Expression of CHD1 & TIF1α is Downregulated upon Differentiation

A. This schematic illustrates the experimental procedure. Epitope tagged-CHD1 mouse ESCs were differentiated into embryoid bodies by shifting them into suspension culture in the absence of LIF. As can be observed in the image (right), after seven days, the embryoid bodies have become large fluid-filled cystic structures.

B. Downregulation of Oct4 and Nanog, two master regulators of pluripotency, was observed by Western blot analysis, confirming successful differentiation of the ESCs. Likewise, induction of Gata4 expression, a marker of endoderm differentiation, was observed. Both CHD1 and TIF1α are preferentially expressed in ESCs, with downregulation occurring upon differentiation.

Our preliminary characterization of TIF1α suggests that it is an interesting protein with potential roles in pluripotency. However, further characterization was hindered by inefficient depletion of TIF1α by interference with short hairpin RNA (shRNA). Moreover, the CHD1-TIF1α interaction appears to only represent a minor

subpopulation. Thus, a judgement call was made to cease further characterization of the CHD1-TIF1α interaction.

3.3 Conclusions

CHD1 is a chromatin remodeller implicated in the maintenance of pluripotency, though the mechanistic details about its role in this process have yet to be determined. To investigate this question, the native CHD1 stable core complex and interactome was purified from mouse ESCs, employing a novel heparin extraction method that is broadly applicable for the solubilization of chromatin-bound proteins. Like its budding yeast and *Drosophila* orthologues, mouse CHD1 probably does not assemble into a stable stoichiometric protein complex that withstands exposure to buffers of high ionic strength.

The low-stringency CHD1 purification was limited by a high level of background contaminants – the silver-stained gels lacked a clear and dominant CHD1 band. Unsurprisingly, even proteins enriched in the CHD1 sample compared to the control had a low signal-to-noise ratio, making it difficult to easily identify specific CHD1 interactors. Nevertheless, one of the mass spectrometry hits, TIF1α, was validated to specifically interact with CHD1. However, the CHD1-TIF1α interaction represents a minor subpopulation of the two proteins. Preferential expression of CHD1 and TIF1α in mouse ESCs was nevertheless noted, suggesting roles in the pluripotent cell state. Although TIF1α deserves further characterization in its own right, the inability to demonstrate a functional connection between CHD1 and TIF1α meant that this project was deemed untenable.

Chapter 4. Results II: Purification of the Endogenous CHD5 Complex from Mouse Tissues

4.1 Aims

Another member of the CHD subfamily of SWI2/SNF2-like chromatin remodelling ATPases, CHD5 is characterized by its tissue-specific expression in neurons and the testis (Zhuang et al., 2014). Unsurprisingly, it is important for the neural development – depletion of CHD5 from the developing mouse neocortex blocks differentiation, resulting in the accumulation of immature neural progenitor cells (Egan et al., 2013). The main phenotype of *CHD5* knockout mice is male infertility due to impaired spermiogenesis, reflecting a role of CHD5 in regulating histone to protamine transition (Li et al., 2014, Zhuang et al., 2014). CHD5 is also the tumour suppressor gene deleted by loss of chromosome *1p36*, which occurs in 70-80% of high-risk neuroblastomas (Bagchi et al., 2007, Fujita et al., 2008).

CHD5 is a homologue of CHD3 and CHD4, sharing 68% and 72% amino acid identity respectively. CHD4 and CHD3 are the catalytic subunits of the ubiquitous NuRD complex (Zhuang et al., 2014). It is suggested that CHD5 forms a neuronal NuRD-like complex as co-immunoprecipitation of CHD5 with other NuRD complex subunits, including HDAC2, RBBP7 and MTA3, has been detected in mouse brains (Potts et al., 2011). Yet, this hypothesis has not been conclusively substantiated by biochemical purification of a stable CHD5 complex. Thus, we decided to purify endogenous CHD5 from mouse brains to clarify the existence of a stable CHD5 complex and its subunit composition. We expected this work to be informative in elucidating the biochemical function and regulation of CHD5, a developmentally important, tissue-specific chromatin remodeller.

4.2 Results

4.2.1 Purification & Mass Spectrometry Analysis of CHD5 Transiently Overexpressed in 293T Cells

As CHD5 is a tissue-specific chromatin remodeller, it is ultimately essential to purify it from an appropriate tissue source – in this case, brains or testes. However, as an initial proof-of-concept experiment, epitope-tagged CHD5 was purified following transient overexpression by calcium phosphate transfection in 293T cells. Despite expression being regulated by a strong CMV promoter, the level of overexpression achieved was initially limited. However, co-transfection with a plasmid expressing the Ebola virus protein, VP35, dramatically enhanced exogenous expression of CHD5 (Figure 4.1A). Co-transfection of mammalian cells with a VP35-encoding plasmid has been previously characterized as a general method of enhancing transient overexpression of target proteins (Gantke et al., 2013). The precise mechanism for this phenomenon remains unclear, but it is believed that the VP35 protein reduces the extent to which double-stranded DNA plasmids activate PKR, a double-stranded RNA-dependent protein kinase that inhibits translation. A striking increase in the level of CHD5 overexpression is observed in the presence of only a small quantity of the VP35 plasmid (a quarter of the amount of CHD5 plasmid transfected) (Figure 4.1A). As a control, 293T cells were transiently co-transfected with plasmids encoding CHD1 and VP35. CHD1 is an appropriate negative control since it is a chromatin remodeller of comparable size and has been established to be monomeric (Figure 3.6B and (Tran et al., 2000, Lusser et al., 2005, Ehrensberger and Kornberg, 2011)).

CHD5 was purified by FLAG affinity chromatography to levels detectable by silver staining (Figure 4.1B). Highly stringent conditions, including both high (750mM sodium chloride) and low (50mM sodium chloride) washes to disrupt ionic and hydrophobic interactions respectively, were used for the purification such that there would be enrichment for the most stable CHD5 interactors. Although some background contamination remained present (refer to the multiple bands visible by silver staining in the control CHD1 purification), the CHD5 purification was associated with unique bands. Thus, both the CHD5 and the control CHD1

puriifications were analysed by tandem LC-MS. The proteins that are most highly enriched in the CHD5 purification are listed in Table 4.1. Notably, this list is dominated by components of the NuRD complex, suggesting that CHD5 forms a NuRD-like complex.

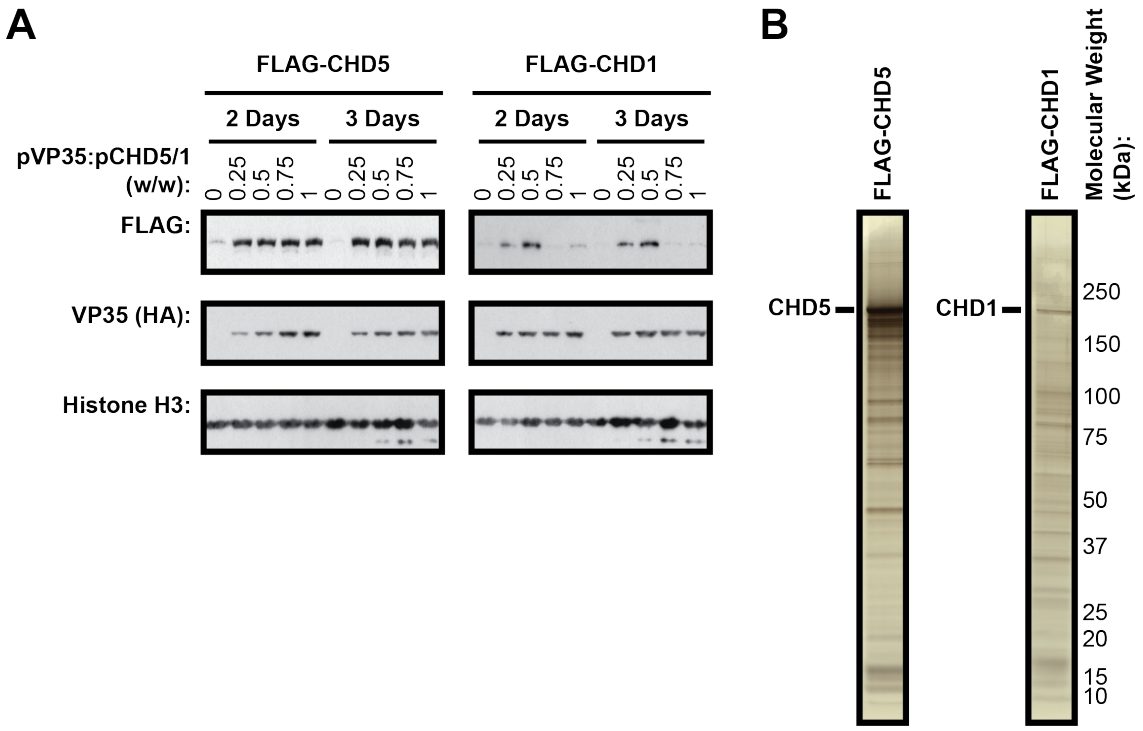


Figure 4.1 – Purification of CHD5 Overexpressed in 293T Cells

A. Expression of CHD1 and CHD5 in 293T cells is dramatically enhanced by co-transfection with a plasmid encoding the Ebola virus VP35 protein.

B. CHD5 was purified by FLAG affinity chromatography under stringent conditions after transient overexpression in 293T cells. CHD1, which is known to be monomeric, was purified in parallel as a negative control. The samples were visualized by silver staining. Although these purifications are associated with some non-specific background contamination (as indicated by the multiple bands in the CHD1 sample), CHD5 appears to specifically enrich for several proteins.

Name	Enrichment	iBAQ Score		Spectral Count		PEP Value
		CHD5	CHD1 Control	CHD5	CHD1 Control	
GATAD2B	5.03	6.03	1	65	0	1.22E-128
PATZ1	4.56	5.56	1	33	0	2.42E-132
MBD3	4.56	5.56	1	13	0	5.52E-51
C5orf55	4.33	5.33	1	3	0	1.49E-02
CHD5	4.19	7.43	3.24	713	3	0
ZKSCAN8	4.16	5.16	1	16	0	3.29E-27
SMN	4.14	5.14	1	2	0	1.49E-04
HOXD13	4.07	5.07	1	10	0	9.94E-26
BAG2	4.01	5.01	1	6	0	9.72E-17
HDAC1	3.96	4.96	1	7	0	1.94E-29

ZBTB5	3.87	4.87	1	9	0	1.20E-49
CDK2AP1;CDK2AP2;DOC-1R	3.84	4.84	1	2	0	4.41E-06
NDUFA4	3.82	4.82	1	1	0	9.17E-08
ADNP	3.80	4.80	1	29	0	2.33E-82
BCL7A;BCL7B;BCL7C	3.79	4.79	1	3	0	1.18E-07
MLF2	3.74	4.74	1	2	0	1.45E-07
MTA1	3.69	4.69	1	12	0	3.25E-52
C3orf67	3.68	4.68	1	1	0	2.75E-04
LTV1	3.67	4.67	1	7	0	1.27E-19
MTA3	3.67	4.67	1	6	0	1.27E-26
MAGED2	3.63	4.63	1	8	0	1.28E-29
FAM101B	3.60	4.60	1	1	0	6.71E-03
FAM207A	3.56	4.56	1	2	0	1.34E-06
STAU2	3.55	4.55	1	7	0	3.38E-21
ECT2	3.53	4.53	1	2	0	1.75E-04
GPATCH4	3.53	4.53	1	3	0	6.40E-17
VPS72	3.51	4.51	1	2	0	6.86E-05
MBD2	3.48	4.48	1	4	0	4.85E-08
AURKB	3.46	4.46	1	2	0	1.14E-07
CBX8	3.44	4.44	1	6	0	1.34E-14

Table 4.1 – The 30 Most Highly-Enriched Proteins by IP of Overexpressed CHD5
The level of enrichment was calculated by normalizing the iBAQ score of a given protein for the CHD5 sample to that for the control CHD1 sample. The thirty most highly enriched proteins are listed in this table. An iBAQ score of 1 (the score of the least abundant protein detected in this sample) was arbitrarily assigned to all proteins without detectable spectra. Most of the NuRD complex subunits (highlighted in yellow) are present in this list of CHD5-enriched proteins.

A notable caveat of this experiment is that CHD5 was transiently overexpressed in a cell type from which it is usually not highly expressed. Hence, it is possible that while CHD5 may be capable of interacting with NuRD complex subunits, this interaction may not actually occur in native tissues because it because it usually interacts instead with an alternative ligand that is of much higher abundance or that has a greater avidity. Nevertheless, this experiment served its purpose as a proof-of-concept study, since it indicates that CHD5 probably forms a stable, multi-subunit protein complex with similarities to the NuRD complex. Consequently, we sought to purify the endogenous, native CHD5 complex from an appropriate tissue source to define its tissue-specific subunit composition.

4.2.2 Purification of Endogenous CHD5 from Mouse Brains

For subsequent purifications of CHD5, a mouse strain expressing epitope-tagged CHD5 mice ($CHD5^{3FTH/3FTH}$) from the native *CHD5* gene locus, generated by our

collaborator, Matthieu Gérard, was used as the source of material. The architecture of the engineered *CHD5* locus is analogous to that described for the epitope-tagged CHD1 ESC line used earlier (Figure 3.1A). The CHD5$^{3FTH/3FTH}$ mouse is indistinguishable from its wild-type counterparts, obtained at Mendelian ratios, and demonstrates no obvious phenotype (personal communication with Matthieu Gérard). Consequently, it is probable that the addition of a C-terminal epitope tag has not affected CHD5 function in this mouse. This mouse is the ideal tissue source for purifications as it combines the unique specificity of an affinity tag with expression of CHD5 at endogenous levels, thereby preserving the native stoichiometries of its interactions. Specifically, brains were used as CHD5 is preferentially expressed in the brain and to a lesser extent, in the testis.

It was noted that the integrity of soluble CHD5 was adversely affected by numerous factors including sonication (Figure 4.2A). For example, when preparing whole cell extracts of SH-SY5Y neuroblastoma cells, sonication significantly reduced the amount of full-length CHD5 in the soluble fraction despite not appearing to affect the vast majority of proteins. Thus, various approaches including nuclease digestion techniques were tested to optimise the extraction of soluble CHD5 in mouse brain extracts. The optimal condition was to combine grinding of snap-frozen mouse brains in liquid nitrogen using a pre-chilled mortar and pestle with high salt (i.e. 750mM sodium chloride) extraction (Figure 4.2B).

A purification protocol featuring affinity and conventional chromatography was developed – purification on an M2 FLAG column with competitive elution using an excess of FLAG peptide was followed by cation exchange chromatography with elution over a salt gradient (Figure 4.3A). The salt elution gradient is particularly useful in identifying *bona fide* subunits of a protein complex since a protein complex should behave as a single molecular entity and consequently, demonstrate perfect co-elution of all of its subunits. Successful enrichment of CHD5 was accomplished by FLAG purification as near-complete depletion of CHD5 from brain extract following passage over the affinity resin was accompanied by efficient peptide elution (Figure 4.3B). Affinity purified CHD5 resolves by SDS-PAGE analysis as several different bands; however, it remains unknown whether these represent distinct isoforms or post-translational modifications.

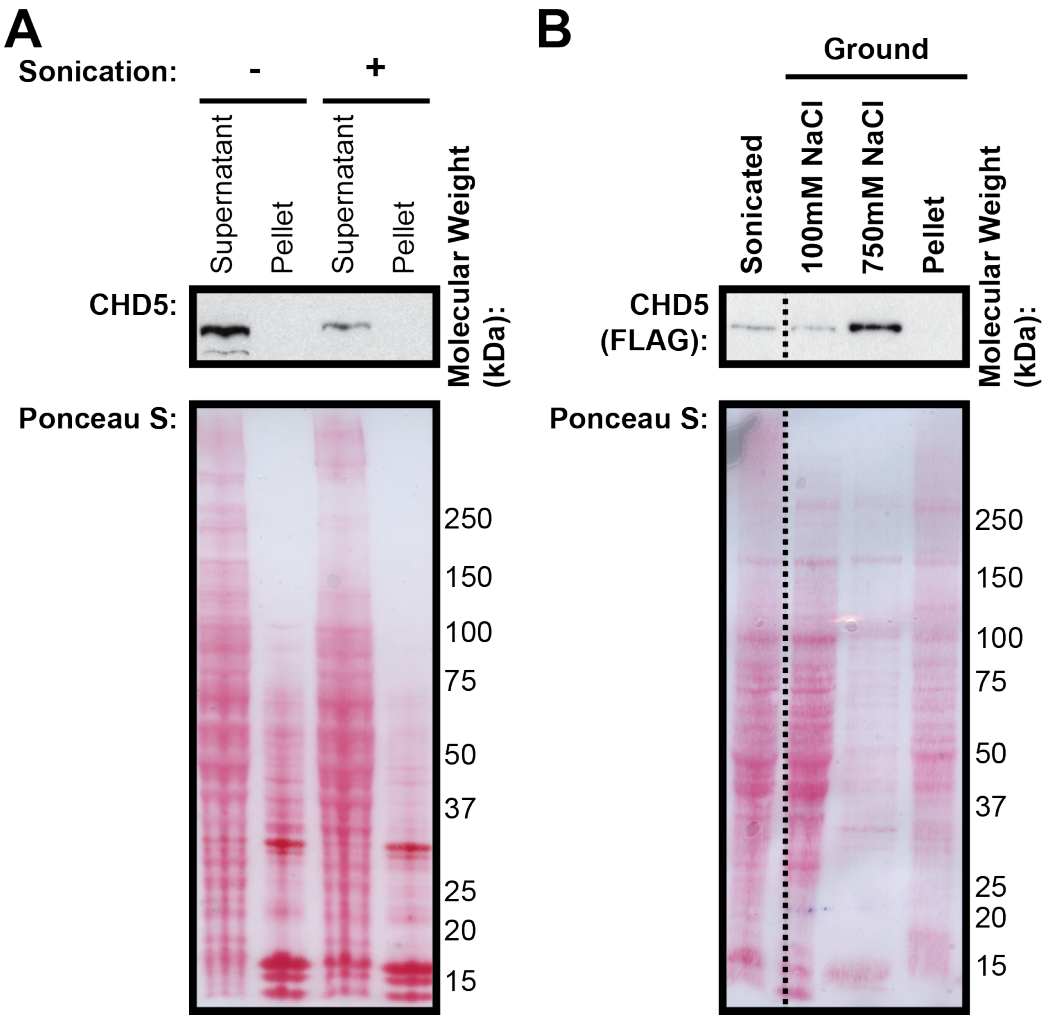


Figure 4.2 – Optimization of Extract Preparation for CHD5 Solubility

A. It was noted that CHD5 is susceptible to sonication-induced degradation. In this experiment, whole cell extracts were prepared from SH-SY5Y neuroblastoma cells with or without sonication. Though sonication did not affect the majority of proteins (Ponceau S), it significantly reduced the level of full-length CHD5 in the soluble cell extract.

B. The level of CHD5 in mouse brain extracts was also adversely affected by sonication (compare lane 1 to 3). Optimal solubilisation of full-length CHD5 was achieved by grinding frozen mouse brains in liquid nitrogen, followed by high-salt extraction.

Significant purification of CHD5 was achieved by cation exchange chromatography since many contaminants are unable to bind to the resin (Figre 4.3C). While CHD5 was purified to a sufficient concentration to be visualized by silver staining, the presence of co-eluting subunits present in stoichiometric amounts was not immediately apparent (Figure 4.3C). Nevertheless, the purification was assessed to be of sufficient quality to warrant tandem LC-MS analysis.

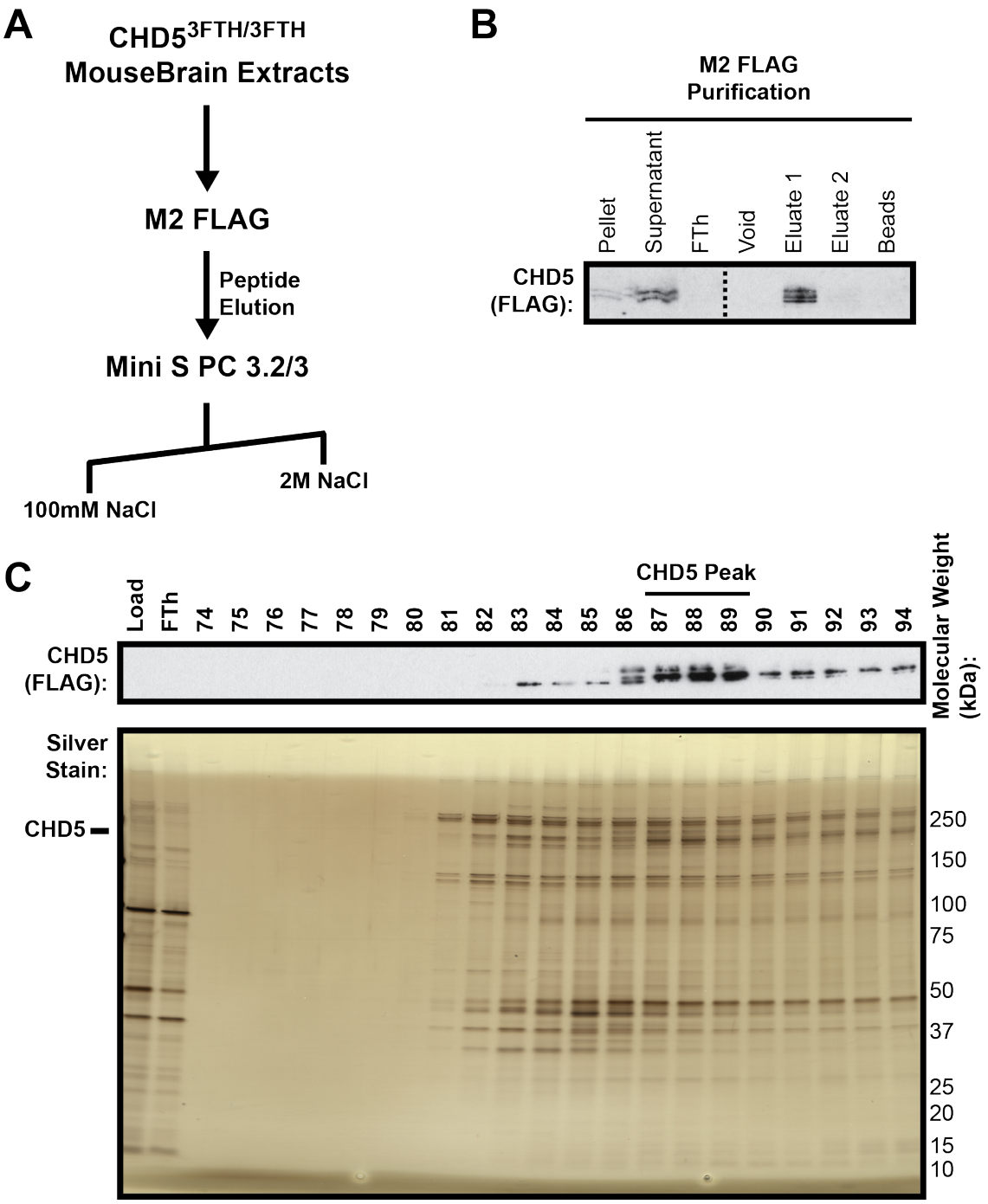


Figure 4.3 – Purification of CHD5 from Mouse Brains

A. CHD5 was purified from mouse brain extracts by sequential FLAG affinity chromatography with competitive elution by FLAG peptide and cation exchange chromatography (Mini S PC 3.2/3). Importantly, the mouse brain extracts have to be prepared by grinding frozen mouse brains with a mortar and pestle followed by high salt extraction.

B. CHD5 is efficiently purified by FLAG affinity chromatography.

C. CHD5 binds efficiently to a cation exchange resin. Some separation of a modified form of CHD5 is achieved in this sharp salt gradient.

Three sets of samples representing different portions of the elution gradient from the ion exchange column were analysed by mass spectrometry. While sample 1 (fractions 87-89) represents the peak fractions of CHD5, sample 2 (fractions 81-83) is effectively a 'negative control' since CHD5 is only present in small quantities in those fractions. Finally, sample 3 (fractions 92-94), comprised of fractions from the tail of the peak, contains intermediate amounts of CHD5 (Figure 4.3C). Each of the three samples were injected on the mass spectrometer three times, allowing for the identification of putative CHD5 subunits by statistical analysis. This was required because of the miniscule amounts of total protein available, and because the CHD5 complex was far from fully purified.

Samples 1 and 2 were directly compared to identify proteins that were specifically enriched in the peak fractions (Table 4.2A). The list of enriched proteins was manually curated to eliminate common non-specific contaminants detected by mass spectrometry. Strikingly, many known subunits of the NuRD complex were identified as being enriched in the peak CHD5 fractions, suggesting that the native CHD5 complex present in mouse brains is similar in composition to the canonical NuRD complex containing CHD3 or CHD4. Nevertheless, several other proteins (highlighted in green) were plausible candidates for being subunits of the CHD5 complex and deserved further validation (Table 4.2A).

The tail of the CHD5 peak can be resolved over a shallower salt gradient as a separate sub-population of CHD5 (Figure 4.5). This separate species of CHD5 roughly corresponds to sample 3 of the initial purification. Comparison of samples 3 and 2 only revealed one protein to be specifically enriched in the second CHD5 peak – KIAA1045 (Table 4.2B).

A. Enriched in Sample 1 vs. Sample 2

Name	Test Score	Label-Free Quantification (LFQ) Intensity									PEP Value
		Sample 1 (87-89)			Sample 2 (81-83)			Sample 3 (92-94)			
		1	2	3	1	2	3	1	2	3	
Spin1	8.76	4.66E+06	3.90E+06	3.07E+06	0	0	0	0	9.46E+05	7.90E+05	8.60E-14
Pura	8.16	1.25E+06	1.27E+06	1.28E+06	0	0	0	1.18E+06	1.16E+06	1.18E+06	1.64E-06
Mta1	7.40	2.06E+06	1.89E+06	1.44E+06	0	0	0	3.15E+06	2.22E+06	1.46E+06	5.58E-15
Khsrp	6.83	1.69E+06	2.00E+06	1.70E+06	3.64E+05	0	4.17E+05	3.92E+06	3.33E+06	3.86E+06	1.26E-25
Csde1	6.53	9.35E	1.23E	1.32E	0	0	0	9.75E	9.54E	1.31E	1.96E-

		+05	+06	+06				+05	+05	+06	10
Elavl2;Elavl4	5.87	8.69E+05	9.33E+05	9.82E+05	2.13E+05	0	0	7.34E+05	6.92E+05	7.07E+05	3.12E-14
Sbf1	5.85	2.86E+07	2.66E+07	3.02E+07	5.44E+06	7.07E+06	5.99E+06	5.74E+06	7.09E+06	5.55E+06	1.04E-76
Chd5	5.26	9.12E+08	8.55E+08	8.43E+08	4.27E+07	5.57E+07	5.47E+07	5.16E+08	4.93E+08	3.91E+08	0
Glyr1	4.94	1.12E+06	1.71E+06	1.60E+06	0	0	0	4.40E+05	1.35E+06	1.09E+06	3.74E-82
Gatad2b	4.91	1.79E+07	2.20E+07	2.27E+07	3.30E+06	2.87E+06	2.48E+06	1.05E+07	7.81E+06	1.05E+07	2.70E-52
Naca	4.87	1.27E+06	1.95E+06	1.70E+06	0	0	0	2.23E+05	3.85E+05	2.71E+05	1.78E-53
Mta3	4.71	4.59E+05	7.77E+05	5.84E+05	0	0	0	8.86E+05	6.67E+05	0	1.44E-12
Hdgfrp3	4.46	2.70E+05	4.61E+05	3.21E+05	0	0	0	3.84E+05	4.04E+05	5.85E+05	4.71E-03
Gatad2a	4.30	3.41E+06	3.46E+06	2.66E+06	7.62E+05	0	0	1.53E+06	1.32E+06	2.02E+06	1.16E-37
Hdac2	4.28	1.71E+06	1.47E+06	6.02E+05	1.31E+05	0	0	1.33E+06	0	0	4.09E-18
Dhx57	3.95	3.03E+07	3.07E+07	2.96E+07	1.49E+07	1.50E+07	1.51E+07	5.31E+07	4.47E+07	2.50E+07	3.71E-107
Api5	3.73	1.22E+05	2.70E+05	2.70E+05	0	0	0	0	0	0	1.10E-06
Rbbp4;Rbbp7	3.45	2.93E+06	4.25E+06	4.04E+06	5.27E+05	4.59E+05	0	5.59E+06	4.49E+06	0	2.31E-57
Mtmr1	3.01	1.62E+06	1.39E+06	1.32E+06	5.14E+05	5.31E+05	4.96E+05	0	4.03E+05	0	9.04E-15
Ddx5	2.78	3.11E+06	4.73E+06	4.88E+06	2.49E+06	2.33E+06	3.07E+06	2.37E+06	2.38E+06	0	1.83E-54
Cdk2ap1;Cdk2ap2	2.62	5.49E+06	4.40E+06	4.99E+06	1.39E+06	1.00E+06	0	2.41E+06	2.41E+06	2.40E+06	1.41E-23
Phxr5	2.49	1.19E+06	4.51E+05	4.54E+05	2.92E+05	0	0	1.43E+05	0	0	1.83E-03
Myef2	2.30	2.07E+05	2.99E+05	1.50E+05	9.09E+04	0	0	7.73E+05	5.98E+05	0	1.92E-09

B. Enriched in Sample 3 vs. Sample 2

Name	Test Score	Sample 1 (87-89)			Sample 2 (81-83)			Sample 3 (92-94)			PEP Value
		1	2	3	1	2	3	1	2	3	
Kiaa1045	3.87	0	0	1.55E+05	0	0	0	2.83E+06	3.24E+06	4.59E+06	7.06E-27

Table 4.2 – Proteins Enriched by Purification of CHD5 from Mouse Brains

A. Each of the three main peaks from the Mini S PC 3.2/3 was injected onto the mass spectrometer three times. Proteins enriched in sample 1 (first CHD5 peak; middle of the gradient) compared to sample 2 (low CHD5; start of the gradient) are listed here, though the data has been manually curated to omit common contaminants. CHD5 is highlighted in blue, whereas the NuRD subunits are coloured yellow. The most interesting of the putative subunits are highlighted in green.

B. This table lists the proteins enriched in sample 3 (second CHD5 peak; end of the gradient) compared to sample 2 (low CHD5, start of the gradient), though the data has been manually curated to omit common contaminants.

The abundance of each protein across the various samples and replicate injections can be clearly visualized in profile plots (Figure 4.4). Hence, it enables assessment of how closely a putative CHD5 subunit recapitulates the elution pattern of CHD5. Unsurprisingly, the NuRD complex subunits behave similarly to CHD5, being detected in fairly high levels in both CHD5 sub-populations (Figure 4.4A). Curiously, the putative CHD5 complex subunits appear to be specific to either one of the two

subsets – Spindlin1, Pura, Sbf1 and Glyr1 are specifically enriched in sample 1, whereas KIAA1045 is enriched in only sample 2 (Figure 4.4B).

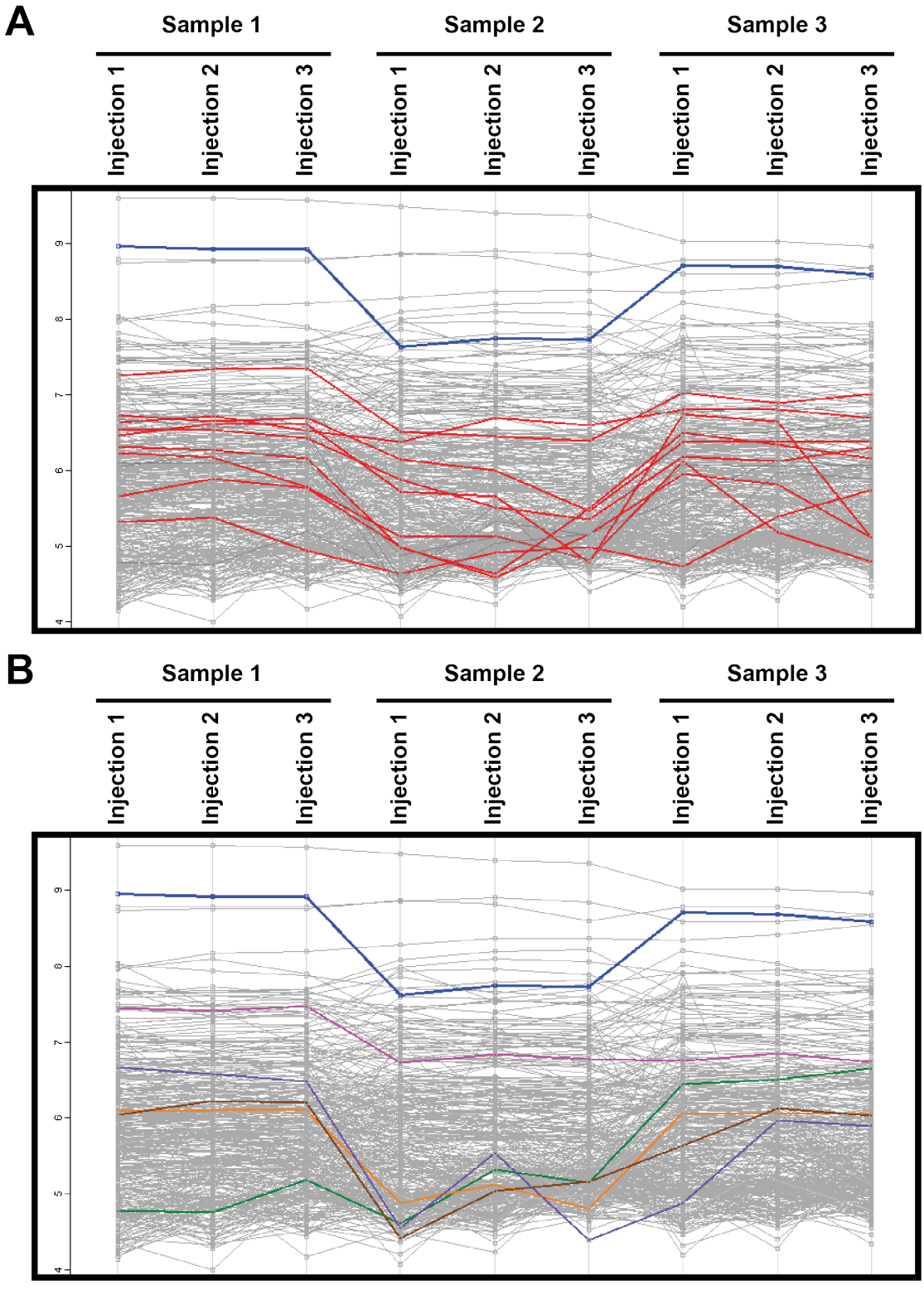


Figure 4.4 – LFQ Intensity Profile Plots of Proteins Identified by Mass Spectrometry from the CHD5 Purification

A. Each grey line represents the LFQ intensities of one protein from the CHD5 purification across the various samples and replicate injections. The CHD5 profile plot is shown in blue, while the NuRD complex subunits are indicated in red. The NuRD complex subunits behave similarly to CHD5 in this purification, being present in relatively high quantities in both CHD sub-populations
B. The profile plots of the putative CHD5 subunits are highlighted here. (KIAA1045 is shown in green, Sbf1 in pink, Pura in orange, Glyr1 in brown and Spindlin in brown.) This diagram illustrates the specificity of these novel factors for one or the other subset of CHD5.

The NuRD complex may undergo combinatorial subunit assembly, particularly in terms of its MBD2/MBD3 and MTA1/MTA2/MTA3 subunits. Thus, the mass spectrometry data was analysed to assess whether the CHD5-NuRD complex differs in composition from the canonical NuRD complex. Both MBD2 and MBD3 are canonical NuRD complex subunits, though their presence in a single complex is mutually exclusive (Le Guezennec et al., 2006). Neither protein was enriched in the CHD5 purification from mouse brains (Table 4.2), though both were associated with over-expressed CHD5 in 293T cells (Table 4.1). Nevertheless, CHD5 has been reported to physically interact with both MBD2 and MBD3 (Potts et al., 2011, Quan et al., 2014, Kolla et al., 2015). This is consistent with the observation that most cell types, apart from ESCs, express both MBD2 and MBD3 (Kaji et al., 2006, Gunther et al., 2013).

MTA1 and MTA3 were identified as being enriched in both the CHD5 purifications from mouse brains and 293T cells (Tables 4.1 and 4.2A). However, MTA2 was not identified to associate with CHD5 in either purification. This is in contrast to previous reports that observed an interaction between CHD5 and MTA2 (Quan et al., 2014, Kolla et al., 2015). Nevertheless, since CHD5 was not purified from mouse brains in sufficient quantities for its stoichiometric complex subunits to be identified conclusively, it is impossible to conclude definitively that MTA2 is absent from the CHD5-NuRD complex in mouse brains. Nevertheless, this observation is certainly intriguing and worthy of further exploration, though the functional significance of a NuRD complex lacking MTA2 is unclear.

4.2.3 Characterization of Putative CHD5 Complex Subunits

Genuine subunits of a stable multi-subunit protein complex should precisely co-elute with each other from a column. Thus, CHD5 was purified again as described earlier (Figure 4.3A), though a shallower salt gradient was used to fully resolve the two CHD5 sub-populations (Figure 4.5). In agreement with the mass spectrometry analysis, MTA1, a subunit of the NuRD complex, is associated with both CHD5 subsets. In addition, out of the five putative subunits identified, Sbf1 and KIAA1045 were also seen to precisely co-elute with CHD5 by Western blotting. Moreover, confirming the mass spectrometry analysis, Sbf1 and KIAA1045 exclusively interact with only one each of the two sub-populations of CHD5 (Figure 4.5).

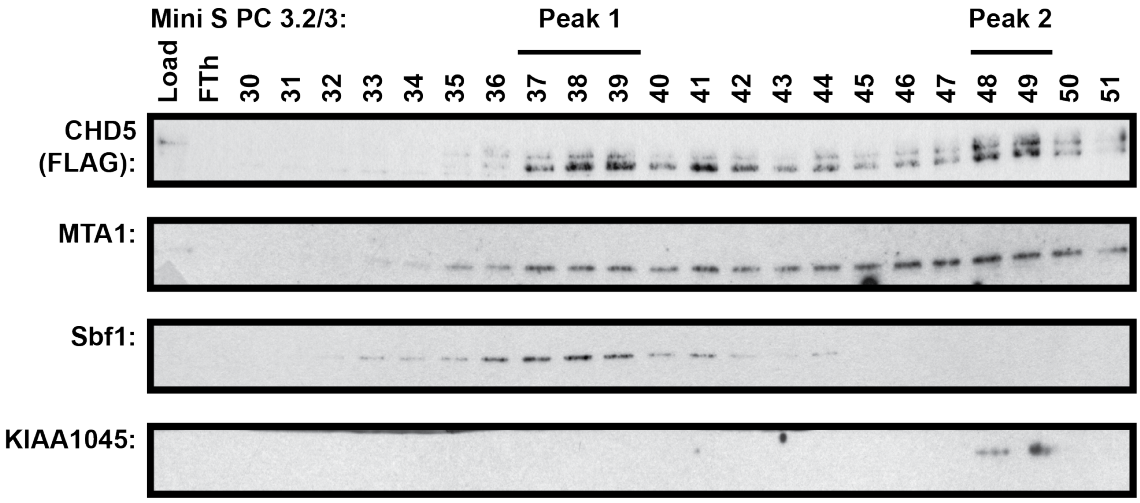


Figure 4.5 – Co-elution of MTA1, Sbf1 & KIAA1045 with CHD5
The CHD5 purification from mouse brains was repeated again, though a shallower salt gradient was used to elute proteins off the Mini S PC 3.2/3 column. This allows the bi-phasic nature of CHD5 elution to be clearly visualized. The NuRD complex subunits, such as MTA1 shown here, appear to co-elute with both subsets of CHD5. However, Sbf1 precisely co-elutes with the earlier-eluting CHD5 peak, whereas KIAA1045 only co-elutes with the second peak. This result may suggest the concomitant existence of multiple CHD5 complexes defined by unique subunits that are exclusively present in one or the other.

Although the purified samples appear fairly clean, it remains possible that Sbf1 and KIAA1045 are non-specific contaminants of the purification. Traditionally, co-elution over multiple columns is required to conclusively identify novel subunits of a stable protein complex (Svejstrup et al., 1995, Zhang et al., 1998). However, such extensive purification of CHD5 is precluded by the inevitable losses associated with each purification step, the limited supply of $CHD5^{3FTH/3FTH}$ mouse brains, and the

low yields of CHD5 due to its relatively low expression levels. Hence, orthogonal approaches were required to validate Sbf1 and KIAA1045. As CHD5 is required for neurogenesis, its expression changes during differentiation of SH-SY5Y neuroblastoma cells. Thus, the changes in expression of CHD5 during neural differentiation was compared to those of Sbf1 and KIAA1045 (Figure 4.6)

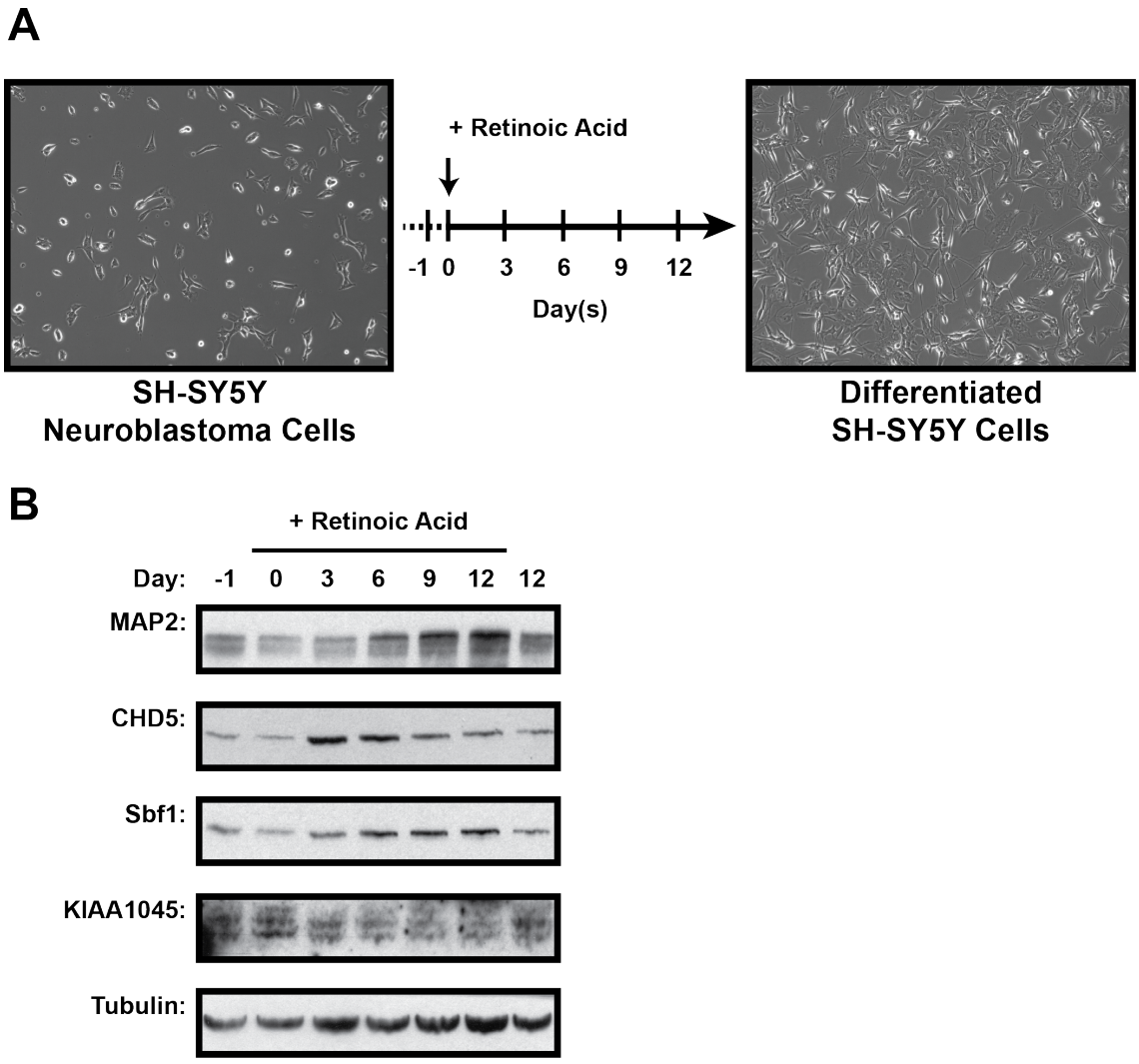


Figure 4.6 – Changes in Expression of Putative CHD5 Subunits upon Differentiation of SH-SY5Y Cells

A. This schematic illustrates the SH-SY5Y differentiation protocol. SH-SY5Y cells seeded in poly-D-lysine-coated dishes were induced to differentiate by 10μM all-trans retinoic acid. Differentiated cells are characterized by the presence of long neurites.

B. Successful neuronal differentiation is indicated by the upregulation of MAP2, a microtubule-associated protein enriched in dendrites. CHD5 levels increased significantly early in the differentiation process before returning to baseline levels. In contrast, Sbf1 levels increased progressively throughout differentiation, while KIAA1045 levels experienced a gradual reduction. Thus, there is no direct correlation between the expression patterns of CHD5, Sbf1 and KIAA1045. As a negative control, a plate of SH-SY5Y cells was also cultured for twelve days, but in the absence of retinoic acid.

The SH-SY5Y neuroblastoma cell line is a well-defined model system for studying neuronal differentiation as it differentiation can be reproducibly achieved by retinoic acid (Korecka et al., 2013, Wang et al., 2014b). Moreover, neuronal differentiation can be monitored by reliable molecular markers such as an up-regulation of the microtubule-associated protein, MAP2, which is enriched in dendrites (Figure 4.6B). In contrast, SH-SY5Y cells cultured for the same duration in the absence of retinoic acid failed to up-regulate MAP2, implying that the changes observed in protein expression were specific to neuronal differentiation and not an artefact of a protracted duration in culture.

CHD5 is expressed at relatively low levels in wild type SH-SY5Y neuroblastoma cells. It is, however, rapidly upregulated early in neuronal differentiation; for example, in this experiment, the highest CHD5 levels were observed three days after inducing differentiation with retinoic acid (Figure 4.6B). While the expression of Sbf1 also increased during neuronal differentiation, it did so with different kinetics to CHD5 – the amount of Sbf1 protein increased progressively over the entire twelve-day differentiation protocol. While KIAA1045 is annotated as being preferentially expressed in brains and in the testis (Genecards), its expression in SH-SY5Y cells changed only subtly upon differentiation, with a slight reduction in its expression being observed (Figure 4.6B). Thus, neither of the two proteins demonstrated to precisely co-elute with CHD5 from the Mini S PC 3.2/3 column displays regulated fluctuations in expression level during differentiation that are similar to that of CHD5. While this result *per se* did not disprove the hypothesis that Sbf1 and KIAA1045 are stoichiometric subunits of stable CHD5 complexes, it was discouraging and indicated that the biochemical data needed to be incontrovertible.

4.2.4 An Alternative CHD5 Purification Strategy from Mouse Brains

In theory, purification over an ion exchange column results in an increase in the concentration of the sample. However, in the case of CHD5, the increase in sample concentration did not materialize, as CHD5 eluted across multiple fractions and was consequently diluted into a relatively large sample volume (Figure 4.5). Nevertheless, it is important to recognize that the considerable width of the CHD5 peak does not reflect a poorly performing column since it is observed reproducibly

with different columns; instead, it suggests that differentially spliced or post-translationally-modified forms of CHD5 are sufficiently distinct from each other in surface charge to be resolved into separate peaks. Regardless of the biological significance of the result, however, subsequent purification of the sample by chromatography is challenging, as there is only a small volume of each CHD5 sub-population. Coupled with the very low protein concentration of the sample and the inevitable losses associated with sample manipulation, additional purification by chromatography would, in all likelihood, reduce the concentration to below the detection threshold. However, we rationalized that ion exchange columns formulated with different resin chemistries would bind non-specifically to different groups of contaminants. Thus, we decided to validate the putative CHD5 complex subunits by assessing whether they co-eluted with CHD5 from a different column.

Consequently, an alternative purification strategy was developed to isolate CHD5 from mouse brains (Figure 4.7A). As before, initial purification by FLAG affinity chromatography successfully enriched for CHD5 from mouse brain extract (Figure 4.7B). A MAbPac SCX-10G cation exchange column (Thermo-Fisher Scientific) was used for the subsequent purification step. While the sulphonic acid functional group of the MAbPac SCX-10G column is similar to that of the Mini S PC 3.2/3 column used previously, it is coupled to a proprietary pellicular substrate that is purported to eliminate hydrophobic interactions due to a hydrophilic surface coating. While elution of CHD5 from the MAbPac SCX-10G was once again in a broad peak, the column appears to separate on the basis of the concomitant presence of a slower-migrating band that is probably indicative of a post-translational modification. Crucially, neither Sbf1, nor KIAA1045, nor any of the other putative subunits previously short-listed, was identified to co-purify with CHD5 by either Western blotting or mass spectrometry analysis (Table 4.3).

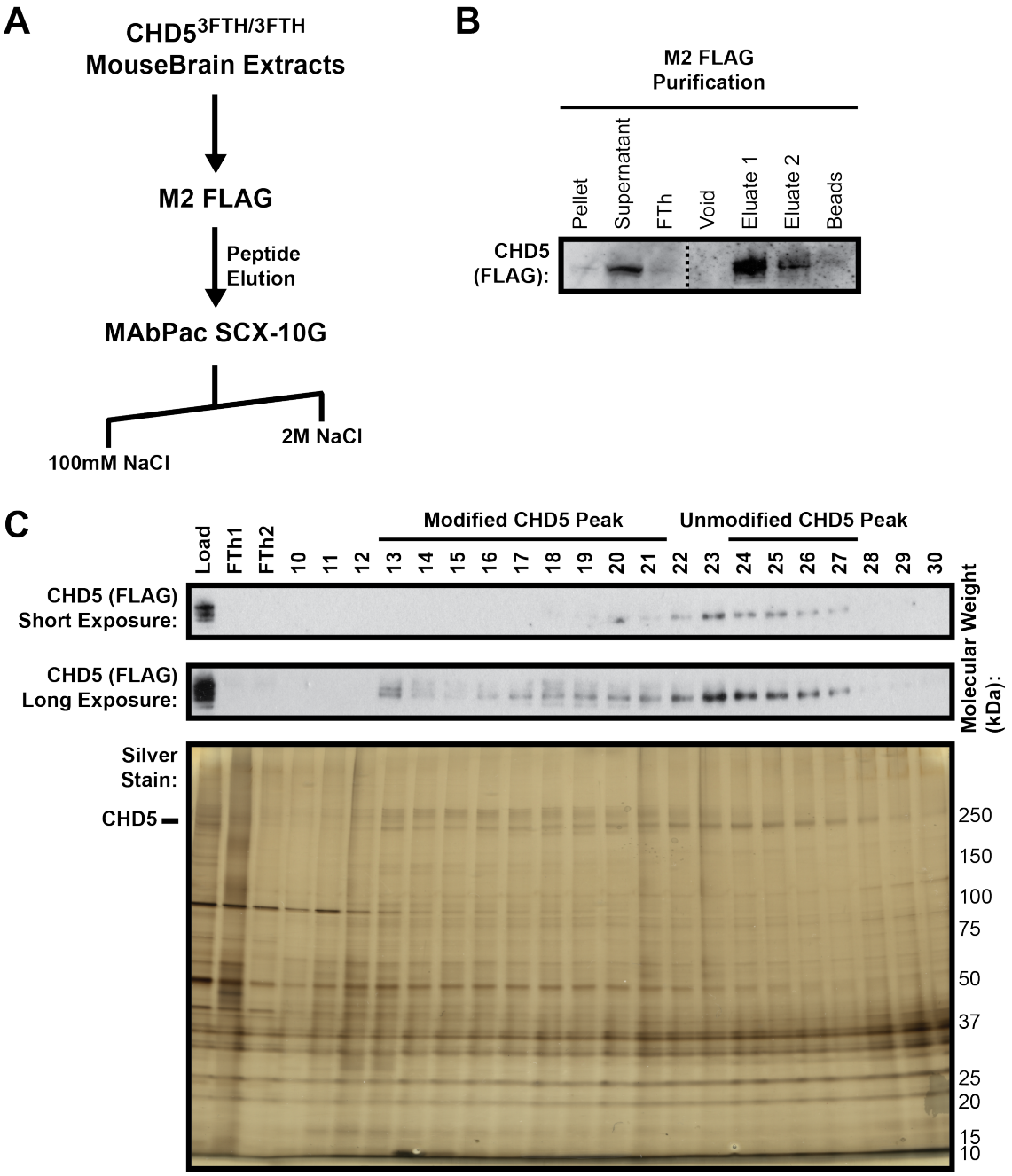


Figure 4.7 – An Alternative Purification Strategy for CHD5 from Mouse Brains

A. CHD5 was purified from mouse brains by sequential FLAG affinity and cation exchange chromatography as before. However, a MAbPAc SCX-10G column was used instead. While the sulphonic acid functional group responsible for binding is similar for both columns, the underlying resin substrate is different. The MAbPac SCX-10G column is supposed to eliminate non-specific hydrophobic interactions and support high efficient separations due to faster mass transfer.

B. As previously observed, CHD5 was efficiently enriched from chromatin extract by FLAG purification.

C. CHD5 elutes across the majority of the salt gradient. However, this column appears to separate CHD5 into predominantly modified or unmodified sub-populations. This is best appreciated on the silver-stained gel.

For the best chance of identifying factors that specifically co-purify with CHD5, all twenty-one fractions from the MAbPac SCX-10G column seen in Figure 4.7C were analysed by tandem LC-MS (Figure 4.8). The fifty proteins that behaved most similarly to CHD5 over all the fractions, using Euclidean distance from the CHD5 reference plot as the ranking criterion, were short-listed (Figure 4.8B). This list of proteins was then manually curated to exclude common contaminants of mass spectrometry experiments (Table 4.3). Strikingly, with the notable exception of NuRD complex subunits (Figure 4.8A), this list is largely free from overlap with the proteins previously identified to co-purify with CHD5 (compare Tables 4.2 and 4.3).

Name	Δ	iBAQ Scores of Fractions																					PEP Value
		10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
Chd5	0	4.0	4.2	4.9	4.6	4.5	4.5	4.4	4.7	4.5	4.6	4.0	4.6	4.5	4.2	4.6	4.4	5.3	4.4	4.1	4.8	4.7	0
Zc3h3	5.2	3.4	3.4	2.8	3.5	3.5	3.2	3.3	3.4	3.6	3.5	3.4	3.3	3.2	3.8	3.4	3.3	4.0	3.1	4.0	3.4	3.8	2.15E-03
Jup	5.2	4.2	3.6	4.0	3.4	3.9	3.7	4.2	3.1	4.3	3.7	2.1	4.2	2.3	2.4	4.1	4.2	3.3	4.0	3.1	3.2	3.9	3.13E-216
Mtap1b	5.7	3.9	4.3	4.6	4.3	4.4	4.1	3.8	4.1	3.9	3.7	3.8	3.4	3.4	3.1	3.3	3.0	2.8	2.7	1.3	2.9	3.1	8.94E-288
Clasp2	6.3	3.4	4.1	5.1	4.8	5.0	4.7	4.3	4.5	4.2	3.9	4.1	3.9	3.7	3.3	3.1	2.9	3.1	1.3	1.3	2.1	3.2	3.06E-273
Anxa2	7.1	3.2	2.8	3.5	3.7	3.6	3.7	3.8	3.4	3.9	3.9	2.5	4.1	3.5	3.2	3.9	3.8	1.3	1.3	1.3	3.4	4.0	3.65E-51
Gatad2b	7.5	3.1	3.8	4.2	4.1	4.2	3.2	3.4	4.0	3.9	3.3	3.5	2.4	2.6	1.3	2.8	1.3	3.7	1.3	2.7	3.0	3.2	1.52E-46
Dsp	7.6	3.7	3.2	2.4	2.6	3.5	3.5	3.9	2.4	4.0	3.4	1.3	3.8	1.6	1.6	3.7	3.7	2.9	3.6	2.3	3.2	3.8	1.95E-197
Hist1h4a	7.8	4.6	3.7	3.6	3.2	3.1	1.3	4.4	3.4	3.7	1.3	3.6	4.6	3.1	1.3	3.4	1.3	4.7	3.9	1.3	4.1	3.9	5.53E-37
Sprr2h	8.0	2.9	3.5	3.5	3.3	3.9	4.1	3.0	3.1	3.9	3.5	1.3	4.3	1.3	1.3	3.4	3.6	3.8	3.9	1.3	1.3	3.5	5.32E-05
Ubc	8.6	3.1	2.7	3.2	2.9	2.5	2.7	2.6	2.5	3.1	2.7	1.3	2.8	1.3	2.0	2.7	3.0	3.2	3.0	2.8	3.0	3.4	6.29E-06
Mbp	8.9	3.2	3.9	4.3	4.2	4.8	3.9	3.8	4.6	4.5	3.7	4.0	1.3	3.3	1.3	2.9	1.3	3.5	1.3	1.3	1.3	1.3	1.98E-70
Aldh1l1	9.7	5.4	6.0	5.7	4.5	4.8	4.6	2.8	4.0	4.2	2.5	3.8	2.7	2.4	1.3	1.3	1.3	1.3	1.3	1.3	2.2	2.4	0
Dsg1b	9.8	2.9	2.9	1.3	2.8	2.8	2.7	2.6	1.3	3.3	1.3	1.3	3.1	1.3	1.3	2.7	3.0	2.9	3.1	3.5	3.6	3.9	3.35E-08
Cacna1d	10.0	2.9	3.0	2.7	3.2	3.0	3.2	3.2	3.2	3.2	3.0	3.3	1.3	3.0	2.9	2.8	2.5	1.3	1.3	1.3	1.3	1.3	2.17E-04
Eef1a1	10.2	3.2	1.3	2.7	2.9	3.0	2.4	3.7	3.1	2.9	1.3	1.3	3.9	1.3	1.3	3.3	3.3	1.3	1.3	1.3	3.6	3.5	5.94E-20
Prdx2	10.3	3.7	3.2	3.2	3.1	3.3	1.3	3.2	2.7	3.6	3.0	2.7	3.7	1.3	1.3	3.6	3.4	1.3	1.3	1.3	1.3	1.3	2.54E-03
Pkp1	10.8	2.7	2.3	1.3	2.3	3.1	2.2	4.2	1.3	3.2	3.1	1.3	3.6	1.3	1.3	3.3	3.0	1.3	1.3	1.3	2.6	3.1	1.59E-56
Tada2b	11.0	3.0	3.2	1.3	3.5	3.0	3.5	1.3	1.3	3.2	3.4	1.3	3.1	1.3	1.3	3.2	3.3	1.3	1.3	1.3	1.3	3.4	2.67E-05
Atp6v0a1	11.4	4.3	4.4	4.8	4.1	4.7	4.1	1.3	3.8	3.9	1.3	3.1	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	4.37E-85
Stk38	11.6	2.3	3.9	5.1	3.6	4.7	3.2	1.3	3.9	3.4	2.7	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	2.33E-105
Atp6v0d1	11.7	4.0	4.1	4.0	3.6	4.0	1.3	1.3	3.7	3.5	2.6	2.7	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.00E-11
Tnrc6b	11.8	2.6	3.4	4.0	3.5	3.6	3.5	2.5	2.2	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	3.3	1.3	1.3	1.3	1.3	0
Dhx57	11.8	1.3	1.3	1.3	1.7	1.9	2.3	2.8	3.2	3.0	3.2	3.2	3.0	2.8	2.1	2.4	1.3	1.3	1.3	1.3	1.3	1.3	3.23E-26
Mta1	11.9	1.3	3.2	3.7	3.3	3.6	3.1	1.3	3.1	3.5	1.3	3.0	1.3	2.2	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.28E-20
Sprr1b	11.9	1.3	1.3	1.3	1.3	3.8	3.5	1.3	1.3	3.7	3.2	1.3	4.2	1.3	1.3	3.1	3.2	1.3	1.3	1.3	1.3	3.3	4.59E-04
Rbbp4	11.9	3.1	3.6	3.5	2.9	3.4	2.5	1.3	3.5	3.5	1.3	3.0	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.16E-07
Hdac2	12.0	1.3	3.5	3.9	2.6	3.6	1.3	1.3	3.4	3.7	1.3	2.5	1.3	1.3	1.3	1.3	1.3	3.2	1.3	1.3	1.3	1.3	1.76E-12
Cdk2ap1	12.1	3.3	4.1	4.5	4.3	4.2	1.3	1.3	4.0	3.9	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.24E-05

Table 4.3 – Proteins Enriched by Purification of CHD5 Using the MAbPac SCX-10G Column that Behave most Similarly to CHD5

Fractions from the MAbPac SCX-10G gradient were analysed by LC-MS. The elution pattern of each protein across the gradient was compared to that of CHD5 on the basis of Euclidean distance (the Δ column in this table). Proteins that were undetected in are sample were arbitrarily assigned the iBAQ score of 1.3, which is the score of the least abundant protein detected in the sample. The 50 proteins that behaved most similarly to CHD5 (indicated in blue) are listed in this table, though the dataset was manually curated to exclude well-known contaminants. This list of proteins is highly enriched in NuRD complex components (highlighted in yellow). None of the other proteins identified can be convincingly described as CHD5 complex subunits.

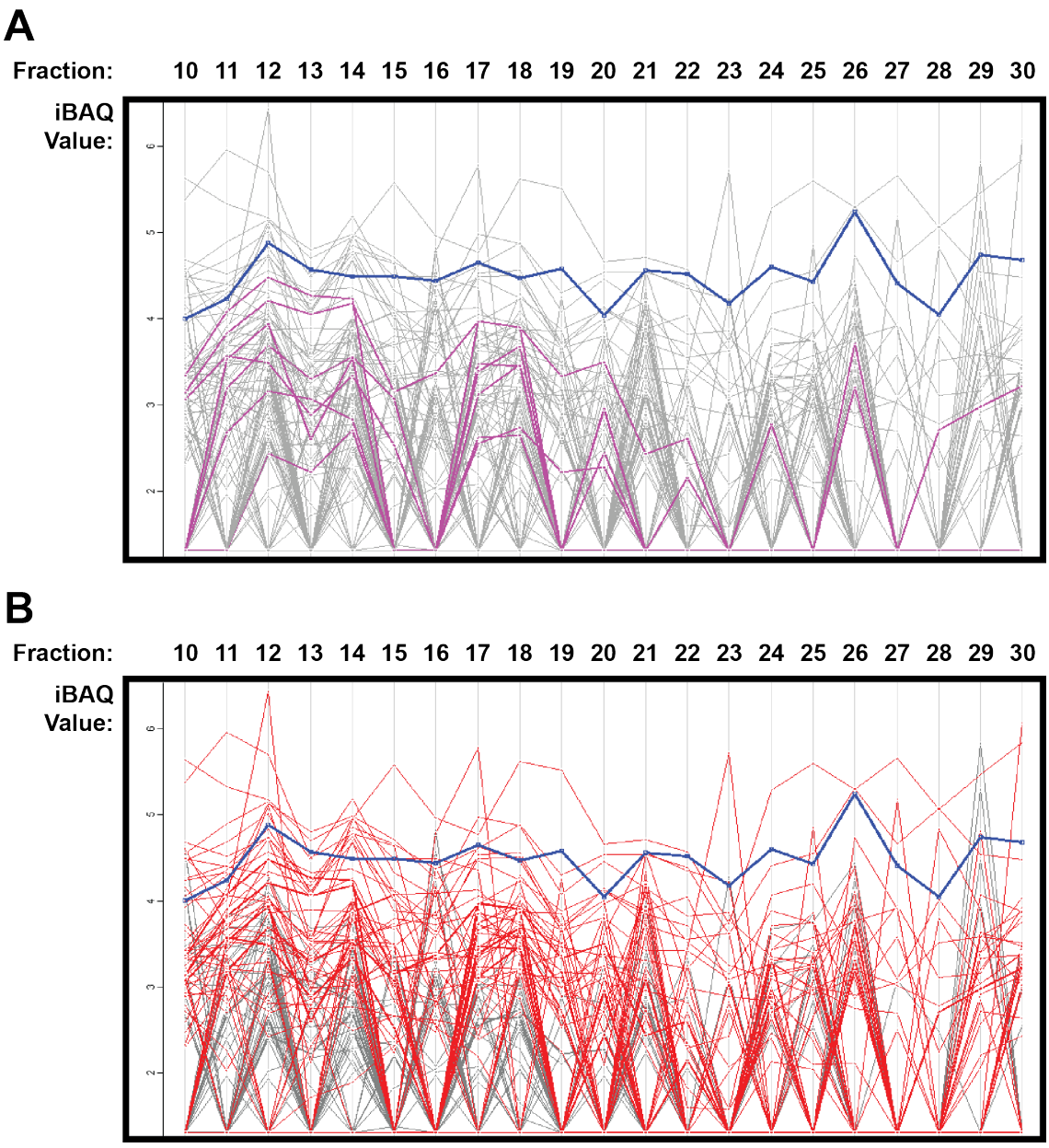


Figure 4.8 – Profile Plots of Proteins Identified from the MAbPac SCX-10G Column Fractions Containing CHD5

A. The abundance (iBAQ values) of CHD5 in these selected fractions from the MAbPac SCX-10G column is illustrated graphically in blue. Each pink line corresponds to a unique NuRD complex subunit. The overall similarity in elution pattern of CHD5 and the NuRD complex subunits can be appreciated.

B. The fifty proteins that behaved most similarly to CHD5, using Euclidean distance from the CHD5 reference plot as the ranking criterion, are indicated in red.

The failure to observe co-purification of Sbf1 and KIAA1045 with CHD5 over the MAbPac SCX-10G column strongly argues against them being genuine subunits a CHD5 complex. Considering all the results holistically, the only strong conclusion that can be drawn is that CHD5 forms a NuRD-like complex in mouse brains.

However, the precise composition of the complex cannot be determined from these data since CHD5 could not be purified from mouse brains in sufficient quantities for all of its stoichiometric subunits to be clearly visualized by silver staining (Figure 4.3C and 4.7C). Unfortunately, this study was unable to identify novel subunits of the CHD5 complex. The lack of overlap between the lists of proteins identified to co-purify with CHD5 using the two complementary purification protocols argues against the existence of such subunits. Given the relatively small quantities of CHD5 and correspondingly, its subunits, that could be purified, it is possible that some subunits were below the detection limit of mass spectrometry. As each tryptic peptide has a certain probability of failing to give rise to detectable spectra, failure of detection by mass spectrometry is more of an issue for small proteins that are inherently limited in the number of peptides that they can generate. Yet, the unparalleled sensitivity of modern mass spectrometers suggests that this is an unlikely possibility, though one that cannot be formally excluded.

4.3 Conclusions

CHD5 is a tissue-specific chromatin remodeller that is specifically expressed in the central nervous system and in the testis. While CHD5 has been previously described to interact with components of the NuRD complex, we anticipated that its complex would contain unique subunits that confer upon it the ability to regulate developmental processes. To address this question, we sought to isolate the native CHD5 complex from mouse brains, a tissue in which CHD5 is preferentially expressed. While an efficient purification strategy combining the specificity of FLAG affinity chromatography and the resolution of ion exchange chromatography was developed, the yields remained modest due to the low expression of CHD5. Although extensive mass spectrometry analysis was performed to thoroughly identify proteins co-purifying with CHD5, no novel complex subunits could be reliably identified for CHD5. However, NuRD complex subunits were consistently enriched in these CHD5 purifications, thereby reinforcing the paradigm of a NuRD-like CHD5 complex.

Yet, the question of how CHD5 achieves its tissue-specific functions, for which CHD3 and CHD4 cannot compensate via functional redundancy, remains an unresolved matter (Egan et al., 2013). Indeed, the failure to identify novel complex subunits specific to CHD5 suggests that the CHD5 protein autonomously confers these unique functions upon itself. Yet, CHD5 shares significant sequence similarity with CHD3 and CHD4, and lacks additional annotated protein domains that might affect its genomic localization or substrate specificity. Thus, additional biochemical and structural studies characterizing the CHD5 protein are required to establish a framework through which its roles in development can be understood mechanistically.

Chapter 5. Discussion I: Insights from the Purification of Endogenous CHD1 and CHD5 from Native Tissues

CHD1 and CHD5 are chromatin remodelling SWI2/SNF2-like ATPases that have important functions in the development of higher eukaryotes. Yet, these enzymes remain poorly characterized at a mechanistic level. Here, the purification of endogenous CHD1 and CHD5 from appropriate tissue sources has been described. However, no novel complex subunits were conclusively identified for either protein. Similar to its budding yeast and fruit fly orthologues, mouse CHD1 does not stably interact with any other co-factor as part of a stoichiometric complex. In contrast, CHD5 appears to assemble into a canonical NuRD complex without any additional unique subunits.

5.1 Combinatorial Subunit Assembly of Chromatin Remodelling Complexes

Chromatin remodellers regulate a diverse range of biological processes; at an organismal level, their functions include the regulation of specific developmental transitions and the maintenance of cellular identity (Lickert et al., 2004, Lessard et al., 2007, Wu et al., 2007, Gaspar-Maia et al., 2009, Ho et al., 2009b). This functional diversity is accomplished through the varied spectrum of specific activities displayed by individual chromatin remodellers, but also through the combinatorial assembly of chromatin remodelling complexes. In general, the core catalytic subunit is shared between the different complexes, while the unique subunits modify its genomic localization or its interaction partners. This represents an elegant, yet general biological mechanism that offers specificity of function without requiring cells to evolve a multitude of specialized chromatin remodelling paralogues. The BAF complex, the mammalian orthologue of the yeast SWI/SNF complex, is undoubtedly the most prolific chromatin remodeller at exploiting combinatorial subunit assembly. For example, unique BAF complexes specific to ESCs, neural progenitor cells, post-mitotic neurons and myocytes are essential for

the maintenance of pluripotency, self-renewal of neural progenitors, dendrite growth, and cardiac development respectively (Lickert et al., 2004, Lessard et al., 2007, Wu et al., 2007, Ho et al., 2009b). In fact, combinatorial subunit assembly has even been exploited in a subset of synovial sarcomas to facilitate carcinogenesis – the SS18-SSX fusion protein, which arises from a specific chromosomal translocation, competitively assembles into the BAF complex, before mediating transcriptional de-repression of the proto-oncogene, Sox2 (Kadoch and Crabtree, 2013).

Cell type-specific functions have been ascribed to both CHD1 and CHD5. Specifically, CHD1 is essential for maintaining the defining traits of self-renewal and pluripotency of ESCs, whereas CHD5 is a tissue-specific chromatin remodeller that is required for the terminal differentiation of neurons (Gaspar-Maia et al., 2009, Egan et al., 2013). Consequently, we speculated that the parallels between the BAF complexes and these CHD proteins extended beyond possessing similar biological functions, to also being organized in a combinatorial manner. However, CHD1 does not exist as part of a stable, multi-subunit protein complex. CHD5 appears to be assembled into the canonical NuRD complex; neither definite preferences for specific MBD or MTA proteins, nor novel complex subunits was identified. Thus, these CHD chromatin remodellers must have acquired functional diversity through a means other than by combinatorial assembly; although the precise mechanistic details remain uncertain, it is possible to construe speculative, yet plausible models for their activities, as discussed below.

5.2 CHD1 Functions as a Monomer

Monomeric chromatin remodellers are not unprecedented. Indeed, the budding yeast and fruit fly orthologues of CHD1 are characterized to be principally monomeric enzymes that can nonetheless regulate transcription termination and pre-mRNA splicing, or mediate the removal of a promoter nucleosome in a gene activator-dependent manner (Alen et al., 2002, Sims et al., 2007, Ehrensberger and Kornberg, 2011). In fact, it is entirely believable that CHD1 acts principally as a monomer since it possesses DNA-binding ability through its C-terminal SANT and

SLIDE domains. Moreover, its general specificity for DNA may be conferred by the ability of its tandem chromodomains to recognize histone tails tri-methylated at lysine 4 (H3K4me3), an active mark (Flanagan et al., 2005, Sims et al., 2005). (Although the binding specificity for the H3K4me3 mark was demonstrated for the chromodomains of human CHD1, the mouse CHD1 chromodomains are virtually identical, except for a few unconserved residues that are substituted with amino acids of the same charge characteristics.) Correspondingly, ChIP-seq of CHD1 in MEFs indicates that it binds just downstream of the transcription start site of genes that are also occupied by RNAPII, thereby allowing it to evict the promoter nucleosome of actively-transcribed genes at a genome-wide level (Skene et al., 2014). This association with active genes is evolutionarily conserved since CHD1 localizes to the de-condensed interband and puff regions of polytene chromsomes (Kelley et al., 1999).

It may be possible that the main functions of CHD1 in ESCs are the same ones it plays in other cells – to maintain the overall nucleosome organization and to globally facilitate transcription by ejecting the promoter proximal nucleosome that would serves as a physical impediment to transcription (Gkikopoulos et al., 2011, Skene et al., 2014). This model may be able to explain the preferential expression of CHD1 in ESCs observed in this study since higher levels of CHD1 may be required to regulate the greater transcriptional activity observed across the entire genome of ESCs (Efroni et al., 2008).

However, it remains impossible to exclude the possibility that CHD1 may acquire functional specificity from transient protein interactions. Indeed, our failure to definitively identify any interaction partners of CHD1 that contribute to its role in pluripotency probably reflects a technical limitation of our study. As compared to stable complex subunits, transient protein interactions are generally weaker and do not interact with every molecule of the protein of interest. Thus, coupled to the relatively low expression of CHD1, the signal-to-noise ratio of CHD1 interactors was low, compromising identification of genuine hits. These technical limitations could be surmounted through quantitative proteomics, for example by the stable isotope labeling by amino acids in culture (SILAC) technique, since the differences in abundance of a protein between samples could then be measured quantitatively

rather than approximated. Future characterization of the CHD1 interactome in ESCs should also introduce differentiated CHD1-tagged cells as an additional control to distinguish between interaction partners of CHD1 that occur specifically in ESCs and those that bind regardless of cell state.

5.3 The CHD5-NuRD Complex

Unlike combinatorial subunit assembly in BAF complexes, where the catalytic subunit remains constant but accessory subunits are swapped to reflect the specific cell type, the NuRD complex appears to substitute its catalytic subunit to reflect the cellular context. CHD5 was observed to associate reliably with canonical subunits of the NuRD complex. Indeed, using different model systems, several groups have now independently reported that CHD5 is incorporated into canonical NuRD complexes (Kolla et al., 2014, Quan et al., 2014). However, while it is clear that CHD5 has the ability to associate with the fully assembled NuRD complex, it remains uncertain whether or not diversification through the formation of partial sub-complexes that only consist of a few of the NuRD complex subunits along with CHD5 occurs simultaneously.

The possibility that the CHD5-NuRD complex might have strong preferences for specific MBD (i.e. MBD2 or MBD3) or MTA (i.e. MTA1, MTA2 or MTA3) subunits was also considered. Although neither MBD2 nor MBD3 was enriched in the CHD5 purification from mouse brains, both were associated with over-expressed CHD5 in 293T cells. The literature also suggests that CHD5 can physically interact with both MBD2 and MBD3 (Potts et al., 2011, Quan et al., 2014, Kolla et al., 2015). This is consistent with the observation that most cell types, apart from ESCs, express both MBD2 and MBD3 (Kaji et al., 2006, Gunther et al., 2013).

MTA1 and MTA3 were identified as being enriched in both the CHD5 purifications from mouse brains and 293T cells (Tables 4.1 and 4.2A). However, MTA2 was not identified to associate with CHD5 in either purification. This is in contrast to previous reports that observed an interaction between CHD5 and MTA2 (Quan et al., 2014, Kolla et al., 2015). Nevertheless, since CHD5 was not purified from

mouse brains in sufficient quantities for its stoichiometric complex subunits to be identified conclusively, it is impossible to conclude definitively that MTA2 is absent from the CHD5-NuRD complex in mouse brains. Nevertheless, this observation is certainly intriguing and worthy of further exploration, though the functional significance of a NuRD complex lacking MTA2 is unclear. Although all three MTA homologues biochemically co-purify with the NuRD complex, there are indications that they may mediate subtly different biological effects {Xue, 1998 #215;Zhang, 1998 #59;Zhang, 1999 #295;Fujita, 2003 #297;Yao, 2003 #296;Le Guezennec, 2006 #238}. For example, though both MTA1 and MTA2 are potent transcriptional repressors, MTA3 performs this role less efficiently (Yao and Yang, 2003). *In vitro* data also suggests that MTA2 enhances the histone deacetylase activity of the NuRD complex; its homologues, MTA1 and MTA3, however, have not been assessed for this capability (Zhang et al., 1999). Finally, in breast glandular epithelium, MTA3, whose expression is dependent on oestrogen signalling, appears to have a specific role in promoting the expression of the epithelial cell-adhesion molecule, E-cadherin. In breast cancer cells, the absence of MTA3 is associated with E-cadherin down-regulation and epithelial-to-mesenchymal cell transition, which promotes tumour invasion. This is because MTA3 acts, probably directly, as a transcriptional repressor of the transcriptional factor, Snail, which inhibits E-cadherin expression (Fujita et al., 2003).

BAF complexes do not readily exchange free subunit molecules for those already incorporated into existing complexes (Lessard et al., 2007, Wu et al., 2007, Kadoch and Crabtree, 2013). Tissue-specific BAF complexes probably arise because cell type-specific subunits are present at comparatively high concentrations and consequently, are favoured for incorporation into BAF complexes over canonical subunits. Thus, the subunit composition of BAF complexes reflects the relative expression level of cell-type specific subunits. For example, coincident with terminal differentiation of neural progenitor cells into post-mitotic neurons, the expression of the progenitor-specific subunits, BAF45a and BAF53a, are dramatically suppressed, with a corresponding induction of the post-mitotic-specific BAF45b and BAF53b subunits. This abrupt *volte-face* reversal in expression establishes a defined transition in the composition and function of the BAF complex, commensurate with the requirements of the cell at either differentiation state

(Lessard et al., 2007). It is possible that the BAF complex represents an anomaly due to its extreme chemical stability – the complex remains stable in a urea concentration of up to two molar (Kadoch and Crabtree, 2013). However, it could also be suggested that the NuRD complex is regulated in a similar fashion since the expression of CHD5 during spermatogenesis anti-correlates with that of CHD3 and CHD4. CHD5 expression is undetectable in germ cells and during early spermatogenesis, but is readily detectable in post-meiotic late round spermatids. Conversely, CHD3 and CHD4 expression is highest in immature germ cells (i.e. in spermatogonia and spermatocytes), but is considerably diminished later in spermatogenesis (Bergs et al., 2014). However, this formal confirmation of this hypothesis awaits further characterization *in vitro* with purified components.

The main implication of this study is, however, that the functional differences between CHD3/CHD4 and CHD5 can probably be attributed to the catalytic subunits themselves. Yet, CHD5 shares a considerable degree of sequence similarity with CHD3 and CHD4, especially in terms of its histone mark reading and ATPase domains. These paralogues differ most considerably in the C-termini, which do not contain any known functional domains. Small protein domains that auto-regulate the catalytic activity of the helicase-like ATPase domain have been identified in other chromatin remodellers; hence, it is possible that comparable regulatory motifs distinguish CHD5 from CHD3 and CHD4, though extensive biochemical characterization of CHD5 will be necessary to investigate the veracity of that prediction (Clapier and Cairns, 2012, Wang et al., 2014a). Additionally, while CHD5 and CHD4 are both able to increase the accessibility of a chromatin template to restriction enzymes, this is achieved by different catalytic activities – CHD4 slides nucleosomes, whereas CHD5 stably unwraps nucleosomal DNA approximately 40-50 base pairs from the end without disrupting the nucleosome (Watson et al., 2012, Quan and Yusufzai, 2014). Nonetheless, additional biochemical and structural studies are required to fully characterize the differences in specific activities between CHD5 and the canonical NuRD complex catalytic subunits, CHD3 and CHD4, and the mechanisms underpinning these differences.

5.4 Limitations of Proteomic Screens

Mass spectrometry has long superseded Edman degradation as the method of choice for the identification of unknown polypeptides. The main limitation of Edman degradation is limited sensitivity. The main limitation of mass spectrometry is that not all peptides 'fly' equally well; correspondingly, not all proteins are easily detected by this approach. On the whole, however, mass spectrometry benefits from remarkable sensitivity, though this introduces a challenge in that many background contaminants are also identified; consequently, specific hits must be distinguished from non-specific ones. The difficulty of identifying *bona fide* co-purifying interaction partners by mass spectrometry is exacerbated when they are only present in the sample at a low abundance, resulting in a low signal to noise ratio. Unfortunately, the proteomic characterizations of the CHD1 interactome and the CHD5 complex were blighted by this issue, which resulted primarily from the small amounts of CHD1 and CHD5 that could be purified from native tissues. Logistical and practical constraints made scaling up the purifications unfeasible.

The selectivity of proteomic screens can be improved considerably by quantitative proteomics, most notably with the SILAC approach. With conventional mass spectrometry, the failure to detect a given protein in a sample does not prove its absence. With SILAC, if the test sample is labelled with a light isotope and the control a heavy isotope or *vice versa*, the difference in abundance of a specific protein between the test and control samples can be quantified more accurately – since the difference in mass between the light and heavy isotopes is known, it is possible to accurately predict the spectral position of a heavy isotope-labelled peptide based on the spectrum of the same peptide that is differentially light isotope-labelled. Thus, if a protein is detected in the test sample, the absence of that same protein from the control sample can be conclusively determined in the context of SILAC-mass spectrometry. While such a technique does not necessarily improve the signal to noise ratio, it allows the relative abundance of proteins to be compared between samples with greater accuracy. Moreover, it is now possible to label whole organisms, including mice, *in vivo* with stable isotopes, thereby making quantitative proteomics feasible even for studies such as that of CHD5, where mouse brains were used as the source material for purification (Zanivan et al.,

2012). Alternatively, the signal to noise ratio could be improved by increasing the yield of protein from the purification. While this simple strategy is undoubtedly effective, it is frequently challenging to purify large quantities of proteins that are endogenously expressed at low levels in native tissues.

A typical immunoprecipitation-mass spectrometry (IP-MS) proteomic screen identified hundreds to thousands of proteins that appear to be specifically enriched over the control. While common contaminants can be discounted, the selection of hits for further characterization is frequently a biased and arbitrary process. Thus, orthogonal functional screens should ideally complement proteomic screens. For example, simultaneously performing a forward genetic screen by RNA interference along with IP-MS would allow the identification of hits that not only interact with the protein-of-interest, but also regulate the same biological pathway. While validation of selected hits with functional assays is typical of IP-MS experiments – and pursued to a certain extent in this work – it is inherently limited by selection bias in terms of the proteins chosen for further validation. In contrast, combining unbiased proteomic and genetic screens grants the greatest opportunity for the identification of novel factors for any given biological process. Indeed, other members of the Svejstrup Lab have successfully applied this “multi-omics” approach to discovering hitherto uncharacterized factors involved in the transcription-related cellular response to DNA damage (personal communication with Stefan Boeing and Laura Williamson).

Given the largely negative nature of the results described in the previous chapters, it was decided to abandon further investigation of the very scarce CHD proteins, and instead to focus on a different project, described below.

Chapter 6. Results III: Comparison of the Enzymatic Activities of SMARCAD1 and the SMARCAD1-KAP1 Complex

6.1 Aims

SMARCAD1 is a predicted chromatin remodelling enzyme on the basis of its conserved SWI2/SNF2-like ATPase domain. It promotes the maintenance of a repressive chromatin environment by antagonizing histone acetylation and by promoting deposition of the H3K9me3 histone mark (Rowbotham et al., 2011). Furthermore, as SMARCAD1 stimulates DNA end resection, it also promotes the repair of DNA double-strand breaks by homologous recombination (Costelloe et al., 2012). Despite these important cellular functions, SMARCAD1 remains uncharacterized mechanistically. Indeed, it has yet to be demonstrated that SMARCAD1 is enzymatically active. Therefore, we wanted to purify recombinant SMARCAD1 to assess some of its enzymatic properties *in vitro*, including its substrate preference.

SMARCAD1 has been demonstrated to form a constitutive, stoichiometric interaction with KAP1, a transcription repressor that is typically associated with heterochromatin (Rowbotham et al., 2011). Chromatin remodellers often feature as integral components of multi-subunit protein complexes – the accessory subunits often either specify additional specificity in the genomic targets of the enzyme or modulate the remodeller's enzymatic activity. Thus, we also wanted to investigate the effect of KAP1 on the enzymatic activity of SMARCAD1.

6.2 Results

6.2.1 Purification of Recombinant SMARCAD1 & KAP1

Prior biochemical characterization of SMARCAD1 has relied upon immunoprecipitation of tagged protein following overexpression in mammalian cells (Rowbotham et al., 2011). While such an approach has allowed the stoichiometric

interaction between SMARCAD1 and KAP1 to be identified, the yields of purified protein that can be obtained from mammalian cells are, inevitably, limited (preliminary data obtained by Hannah Williams). As detailed biochemical characterization of the enzymatic properties of SMARCAD1 would be most feasible with an abundance of purified protein, we investigated the prospect of expressing it recombinantly in *E. coli*. While an ideal expression system in many regards, obtaining soluble, full-length protein is a frequent challenge when expressing large proteins such as SMARCAD1 in *E. coli*.

To maximise the chances of obtaining soluble, full-length protein, the human SMARCAD1 cDNA was cloned into the pET28a-SUMO vector, allowing SMARCAD1 to be expressed as a fusion protein with an N-terminal hexa-histidine tag and SUMO moiety (Figure 6.1). Compared to other conventional tags, the SUMO tag has been demonstrated to be the most effective at enhancing the solubility of exogenous proteins expressed in *E. coli* (Marblestone et al., 2006). Additionally, the yeast SUMO (Smt3) protease, Ulp1, which can be purified in large quantities, can then be used to cleave the fusion protein immediately after the tandem glycine residues of the SUMO moiety, resulting in recombinant protein with a native N-terminus.

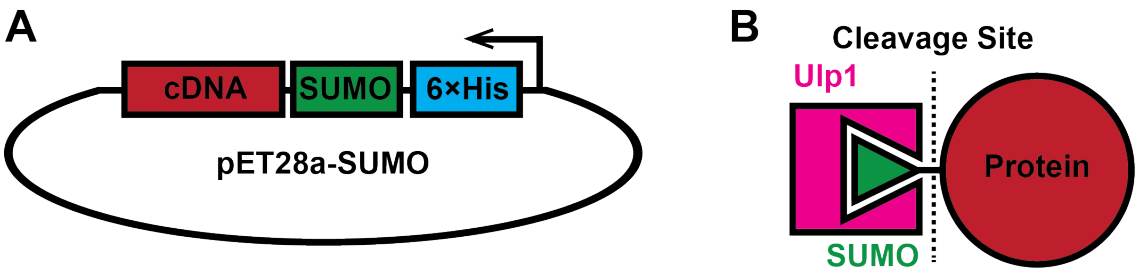


Figure 6.1– The pET28a-SUMO Expression System

A. A schematic depicting the pET28a-SUMO vector used to express most of the recombinant proteins in this study. The hexa-histidine tag at the N-terminus of the fusion protein allows affinity purification by Ni-NTA chromatography, while the SUMO moiety enhances the solubility of many recombinant proteins expressed in *E. coli*.

B. The SUMO-fusion protein expressed from the pET28a-SUMO vector can be converted into its native form by cleavage with the yeast SUMO (Smt3) protease, Ulp1, which cuts specifically after the di-glycine residues at the C-terminus of the SUMO moiety.

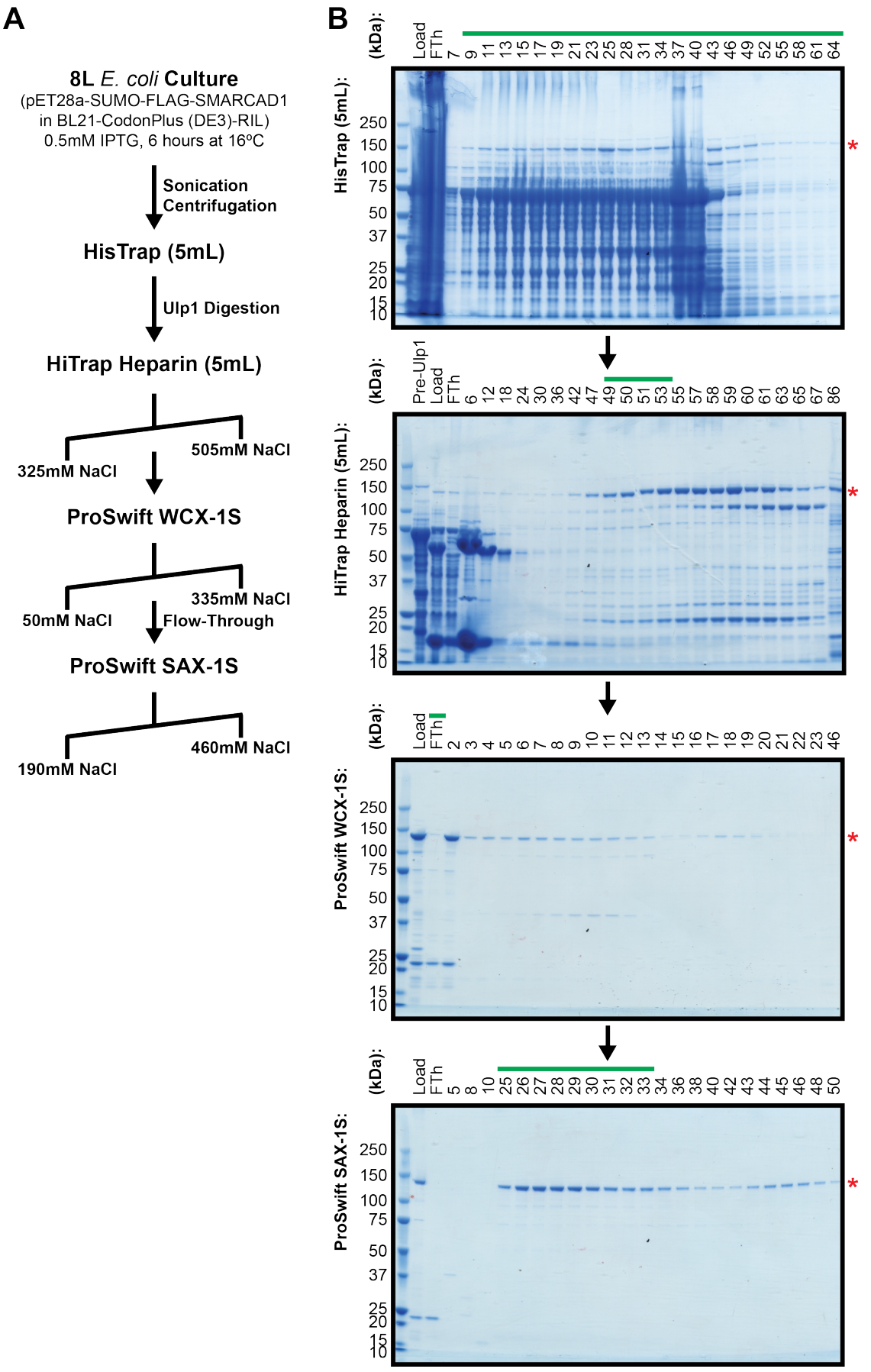

A
8L E. coli Culture
(pET28a-SUMO-FLAG-SMARCAD1
in BL21-CodonPlus (DE3)-RIL)
0.5mM IPTG, 6 hours at 16ºC
Sonication
Centrifugation
HisTrap (5mL)
Ulp1 Digestion
HiTrap Heparin (5mL)
325mM NaCl
505mM NaCl
ProSwift WCX-1S
50mM NaCl
335mM NaCl
Flow-Through
ProSwift SAX-1S
190mM NaCl
460mM NaCl
B
(kDa):
HisTrap (5mL):
Load
FTh
HiTrap Heparin (5mL):
Pre-Ulp1
ProSwift WCX-1S:
ProSwift SAX-1S:
250
150
100
75
50
37
25
20
15
10
*

Figure 6.2 – Purification of Recombinant SMARCAD1

A. This diagram outlines the chromatographic steps required to obtain pure recombinant SMARCAD1 expressed in *E. coli*.

B. These gels, stained by InstantBlue, from a representative SMARCAD1 purification illustrate the progressive increase in purity over the purification protocol. The red asterisks indicate full-length SMARCAD1. The green bars above each gel indicate the fractions that were pooled and loaded onto the next column. The key chromatographic step is HiTrap Heparin chromatography (second gel from the top), which allows the separation of full-length SMARCAD1 from lower molecular weight SMARCAD1 fragments.

Expression trials revealed that SMARCAD1 expression was maximal when cultured at 16°C for six hours following IPTG induction. Pure, recombinant SMARCAD1 was obtained after extensive purification over four columns (Figure 6.2). The most important step was the HiTrap Heparin chromatography as it allowed separation of full-length SMARCAD1 from lower molecular weight SMARCAD1 fragments. Since SMARCAD1 was first purified via an N-terminal hexa-histidine tag, the shorter fragments probably arise from prematurely terminated transcripts or by degradation of the C-terminus. In contrast, the third ProSwift WCX-1S column only achieved marginal improvements in purity and its omission only resulted in slightly more contamination that could be overcome by greater selectivity in pooling fractions.

In contrast to SMARCAD1, KAP1 has been previously purified following expression in *E. coli* (Liang et al., 2011). Indeed, recombinant KAP1 was soluble and could be purified to a high level of purity by a combination of affinity and conventional chromatographic techniques (Figure 6.3). While the main constituent of the purified sample was full-length KAP1, a small quantity of lower molecular weight KAP1 fragments persists. Interestingly, these KAP1 fragments could not be separated from the full-length protein by either anion exchange or size exclusion chromatography (Figure 6.3B respectively). KAP1 has previously been reported to trimerize due to its coiled-coil domain (Peng et al., 2000). Thus, it is possible that a truncated KAP1 molecule that nevertheless retains its N-terminal RBCC domain may form a stable heterotrimer with full-length KAP1. This KAP1 heterotrimer may not differ significantly from a KAP1 homotrimer in terms of surface charge or hydrodynamic radius, potentially explaining the persistence of lower molecular weight KAP1 fragments even after multiple chromatographic steps. The yields of purified recombinant KAP1 from *E. coli* were sufficiently high to permit screening in crystallization trials, though no crystals were ultimately obtained.

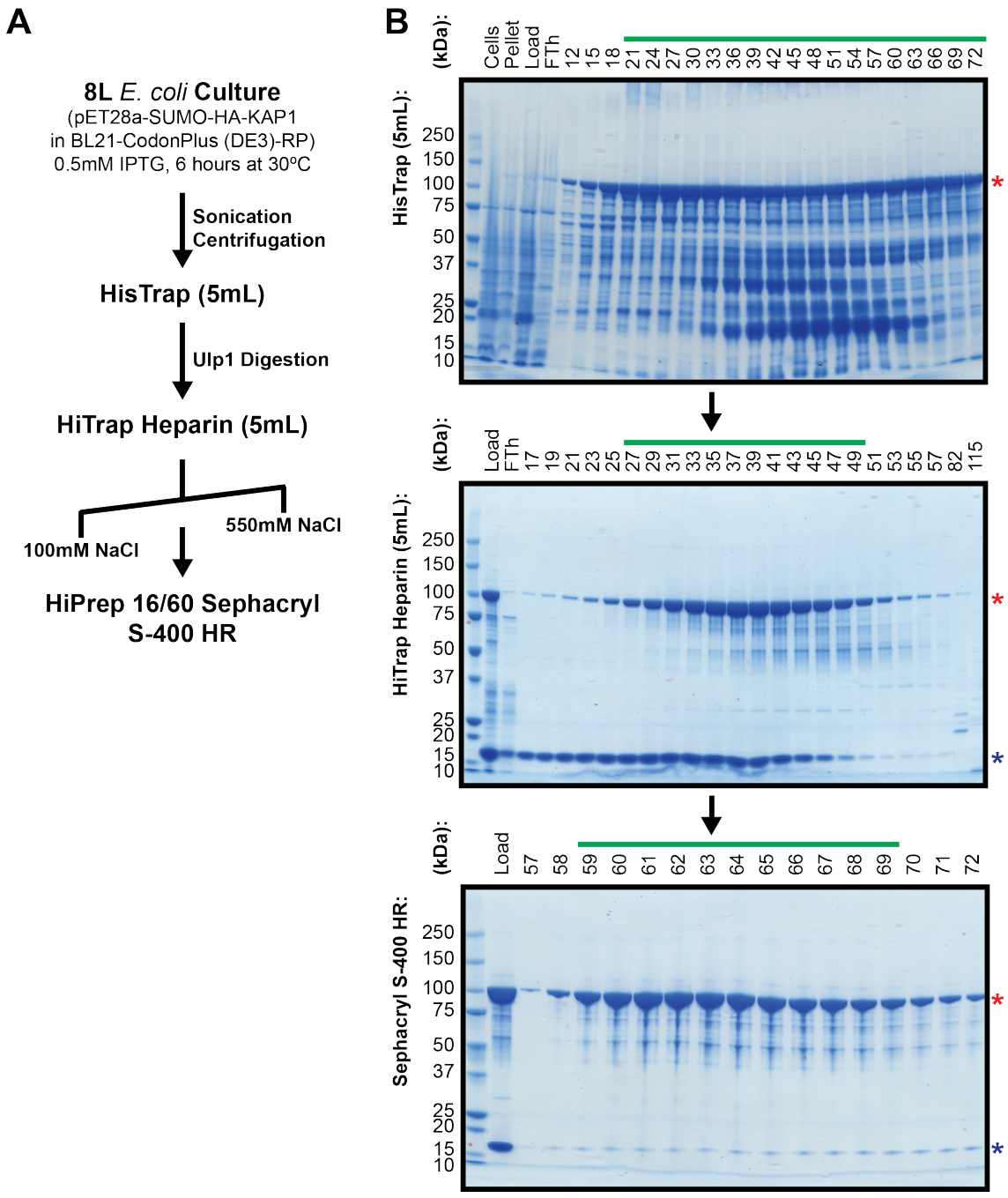


Figure 6.3 – Purification of Recombinant KAP1

A. This diagram outlines the chromatographic steps required to obtain pure recombinant KAP1 following expression in *E. coli*. The final gel filtration chromatography step can be replaced with anion exchange chromatography (e.g. Mono Q 5/50 GL) without affecting either the purity or yield.

B. These gels, stained by InstantBlue, from a representative KAP1 purification illustrate the progressive increase in purity over the purification. The red asterisks indicate full-length KAP1; the blue asterisks denote the cleaved SUMO tag. The green bars above each gel indicate the fractions that were pooled and loaded onto the next column. The lower molecular weight KAP1 fragments cannot be separated from full-length KAP1 by either gel filtration chromatography, probably because these fragments possess an intact coiled-coil domain, thereby retaining the ability to oligomerize with the full-length protein.

6.2.2 Reconstitution of the SMARCAD1-KAP1 Complex

Having devised a protocol to purify recombinant SMARCAD1 and KAP1, the next goal was to reconstitute the SMARCAD1-KAP1 complex *in vitro*. A simple strategy of sequential FLAG and HA affinity purifications on a mixture of FLAG-tagged SMARCAD1 and HA-tagged KAP1 proteins was used (Figure 6.4). However, the recovery of protein following the reconstitution process was poor because of low efficiency binding to the HA column. Further losses were incurred from incomplete elution of the bound complex from the affinity resins by competition with FLAG or HA peptide. Correspondingly, the concentration of the SMARCAD1-KAP1 complex was significantly lower than that of either of the individual components. Nonetheless, the reconstituted SMARCAD1-KAP1 complex was pure and appeared to contain stoichiometric quantities of both SMARCAD1 and KAP1.

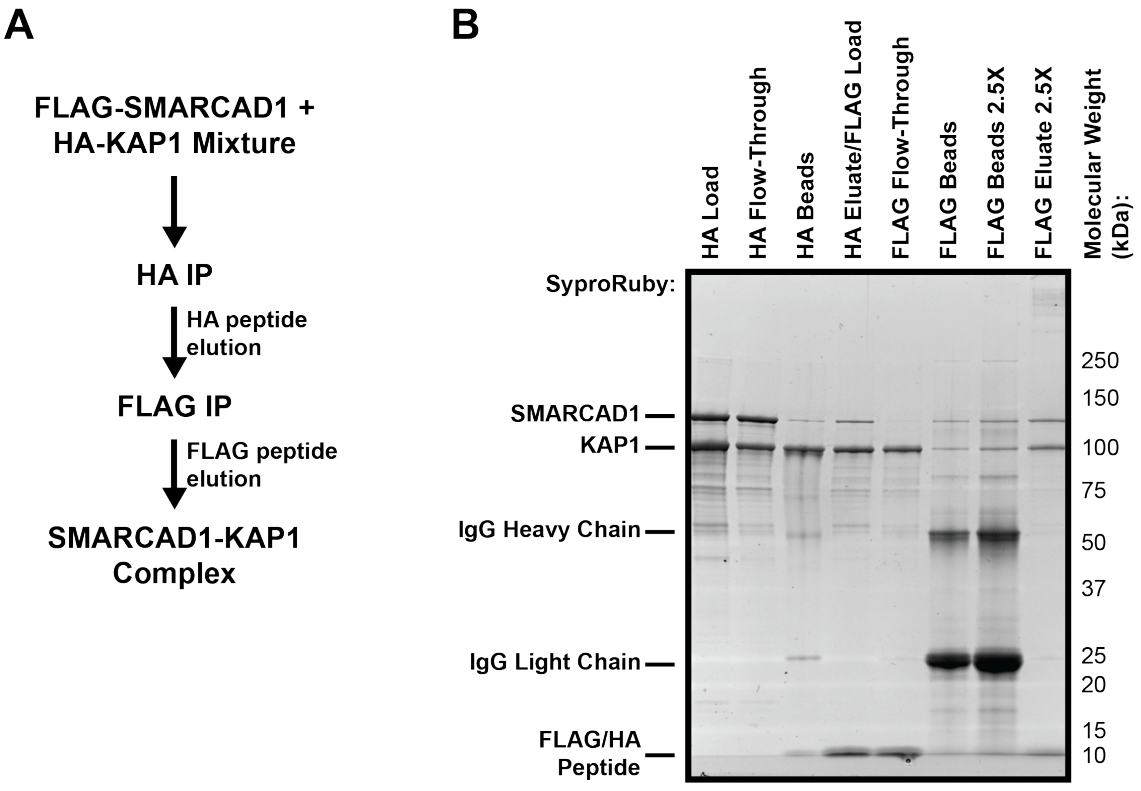


Figure 6.4 – Reconstitution of the SMARCAD1-KAP1 Complex

A. This diagram depicts the sequential affinity purifications performed to reconstitute the SMARCAD1-KAP1 complex *in vitro*.

B. This SyproRuby-stained gel from a representative SMARCAD1-KAP1 reconstitution illustrates the losses incurred at various stages of the tandem affinity purification protocol. Notably, recovery of protein following peptide elution from the two affinity resins was particularly poor. Nonetheless, the final reconstituted SMARCAD1-KAP1 is pure and is comprised of stoichiometric quantities of each protein.

To confirm that the reconstituted SMARCAD1-KAP1 complex behaved as a single molecular species, it was analysed by gel filtration chromatography, along with purified recombinant SMARCAD1 and KAP1 (Figure 6.5). All three samples tested (i.e. SMARCAD1, KAP1 and the SMARCAD1-KAP1 complex) eluted as single peaks from the MAbPac SEC-1 gel filtration column, suggesting mono-dispersity. Crucially, both components of the reconstituted complex precisely co-eluted with each other. Additionally, the complex eluted earlier in the gradient than either SMARCAD1 or KAP1 alone, indicating that the hydrodynamic radius of the complex is larger than that of either of its constituent parts (Figure 6.5B). Taken together, this data suggests that a *bona fide* SMARCAD1-KAP1 complex can be reconstituted *in vitro* from purified components.

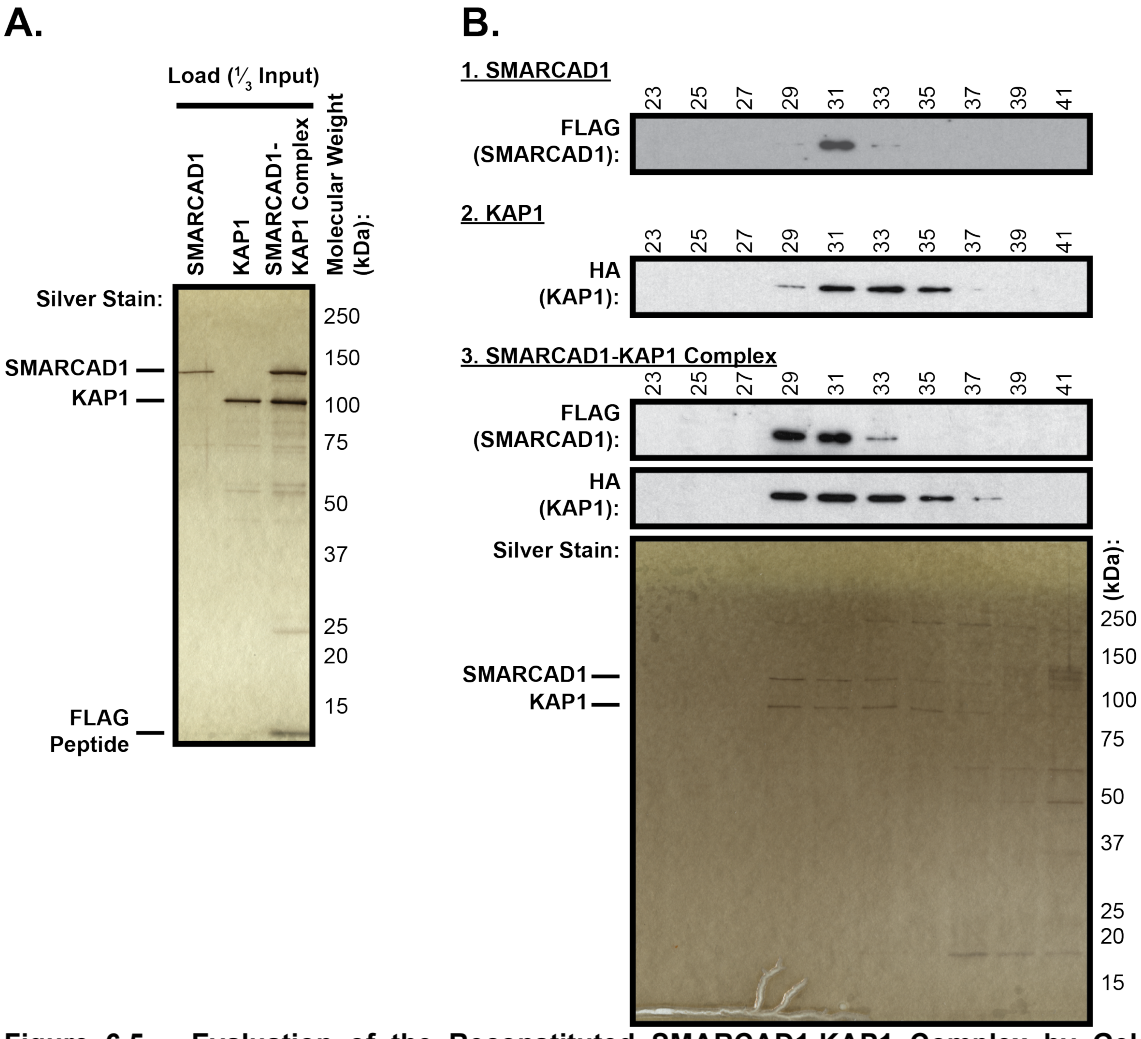


Figure 6.5 – Evaluation of the Reconstituted SMARCAD1-KAP1 Complex by Gel Filtration Chromatography

A. This silver-stained gel compares the samples analysed by gel filtration chromatography; three times the amount of protein present on the gel was actually loaded onto the MAbPac SEC-1 column.
B. Pure SMARCAD1 and KAP1 elute as single peaks from the size exclusion column after the void volumne, suggesting monodispersity of the samples. When the SMARCAD1-KAP1 complex is resolved by gel filtration chromatography, both SMARCAD1 and KAP1 precisely co-elute as assessed by both Western blotting and silver staining, suggesting the formation of a *bona fide* complex. Importantly, the components of the complex elute earlier in the gradient than their individual components (compare the blots in section 3 to those in sections 1 and 2), suggesting that the complex has a larger hydrodynamic radius than either of its constituent parts.

6.2.3 Substrate Preferences of SMARCAD1

Although the enzymatic properties of SMARCAD1 have yet to be described, a reasonable prediction is that it would behave similarly to other chromatin remodellers – displaying ATPase activity that is preferentially stimulated by nucleosomes (Corona et al., 1999). This hypothesis was directly tested in an ATPase assay, where a defined concentration of SMARCAD1 (120nM) was stimulated with limiting concentrations (either 40nM or 80nM) of single-stranded DNA, double-stranded DNA (dsDNA), or reconstituted Xenopus nucleosomes (Figure 6.6). (Although the ATPase assays shown here are representative examples, the results were obtained reproducibly.) SMARCAD1 displayed a clear substrate preference for nucleosomes though dsDNA elicited some ATP hydrolysis (Figure 6.6A & B).

Akin to SMARCAD1, the SMARCAD1-KAP1 complex displayed greater ATP hydrolysis when stimulated by nucleosomes as compared to either ssDNA or dsDNA (Figure 6.6A & B). Notably, however, when stimulated with the same concentration of nucleosome, the SMARCAD1-KAP1 complex hydrolysed considerably less ATP than SMARCAD1, suggesting that KAP1 acts to inhibit the enzymatic activity of SMARCAD1. As a control, equal amounts of the samples of SMARCAD1 and the SMARCAD1-KAP1 complex used in this experiment were resolved by SDS-PAGE and silver-stained, confirming that the concentration of the samples were comparable (Figure 6.6C). Thus, while this possibility cannot be formally discounted, it is unlikely that these differences in activity can be attributed to vastly different concentrations of enzyme being used in these reactions.

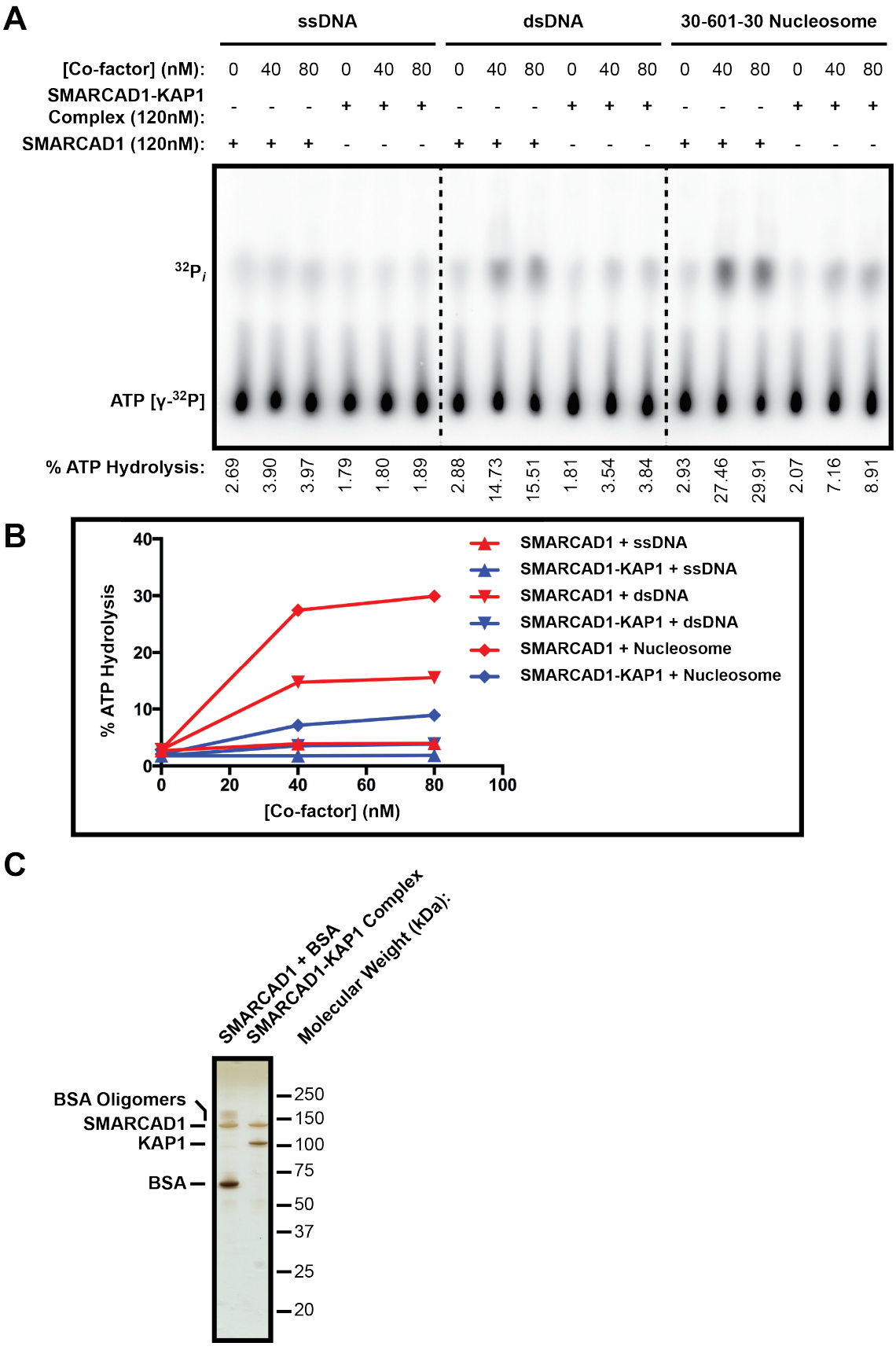


Figure 6.6 – Substrate Preferences of SMARCAD1 & the SMARCAD1-KAP1 Complex

A. In this ATPase assay, the substrate preferences of SMARCAD1 and the SMARCAD1-KAP1 complex were compared by mixing limiting concentrations (40nM or 80nM) of ssDNA, dsDNA, or reconstituted Xenopus nucleosomes with a constant concentration of enzyme (120nM) along with an excess of ATP. Thin layer chromatography allowed free phosphate to be resolved from unhydrolysed ATP; since some radiolabelled ATP [γ-^{32}P] was included in the reaction, the extent of ATP hydrolysis could be determined. SMARCAD1 displays a clear preference for nucleosomes, though double-stranded DNA elicits some ATP hydrolysis. Notably, as shown in this representative experiment, the nucleosome-stimulated ATPase activity of the SMARCAD1-KAP1 complex was considerably lower than that of the SMARCAD1 enzyme by itself. Though this image illustrates the result from a 20-minute reaction, a longer reaction (i.e. 60 minutes) produces a comparable result.

B. The percentage of ATP hydrolysis observed under each condition is plotted in this graph. The red lines relate to reactions with SMARCAD1, while the blue lines relate to reactions with the SMARCAD1-KAP1 complex. The different shapes indicate the substrate added to the reaction. This graph, once again, illustrates the preference of SMARCAD1 for nucleosomes.

C. As assessed in this silver-stained gel, comparable concentrations of SMARCAD1 protein and the SMARCAD1-KAP1 complex were used for these ATPase reactions.

The silver-stained gels (Figures 6.6C, 6.8C, 6.9C & 6.10C) comparing the SMARCAD1 and SMARCAD1-KAP1 samples used in the ATPase assays give the impression that BSA was only added to the SMARCAD1 reactions. In fact, an excess of BSA was present in all ATPase reactions as it was a major constituent of the reaction buffer (***2.1.6.3***). Prior to use in the ATPase assay reactions, the purified SMARCAD1 protein was diluted to a similar concentration as the SMARCAD1-KAP1 complex. BSA was added to the SMARCAD1 sample when it was diluted since in its absence, considerable losses of SMARCAD1 were associated with low sample concentration – it is plausible that the microcentrifuge tubes used have a certain capacity for binding non-specifically to SMARCAD1 molecules that translates into a noticeable problem when the sample concentration is low. While the presence of BSA in the SMARCAD1 sample increased the concentration of BSA in the reaction, the increase was slight in relation to the amount of BSA added to all reactions through the reaction buffer. Moreover, it was also empirically determined the slight differences in BSA concentration caused by the presence of BSA in the SMARCAD1 sample do not affect the ATP hydrolysis activity of either SMARCAD1 or the SMARCAD1-KAP1 complex.

As this is the first instance in which an enzymatic activity can be directly attributed to SMARCAD1, a series of ATPase reactions were set up to determine the optimal conditions for its activity. The two main variables tested were the concentrations of

sodium chloride (NaCl) and magnesium chloride ($MgCl_2$) in the reactions (Figure 6.7). It is clear that the ATPase activity of SMARCAD1 is exquisitely sensitive to the concentration of NaCl – increasing it beyond an optimal concentration of 40mM results in dramatic reductions in activity. In contrast, the magnesium concentration does not appear to drastically affect activity though SMARCAD1 probably has a slight preference for a concentration of 0.1mM $MgCl_2$ (Figure 6.7A & B). On the basis of this knowledge, the salt concentration of all subsequent series of reactions was kept strictly controlled; indeed, to ensure that the concentration of NaCl was minimised and in the optimal range, the only source of salt in these reactions was from the purified protein samples. While it is known that some chromatin remodellers (e.g. RSC) are salt-sensitive and display a preference for potassium acetate compared to sodium chloride, this was not a hypothesis that could be easily tested here, as it required all the enzymes used in this study to be re-purified under different conditions.

Finally, to exclude the possibility that the nucleosome-stimulated ATPase activity observed with SMARCAD1 or the SMARCAD1-KAP1 complex was actually due to a minor contaminant, the 'ATPase-dead' K528A SMARCAD1 mutant was expressed, purified in the same fashion, and tested in an ATPase reaction. Importantly, no ATP hydrolysis over background levels was associated with the K528A mutant, as would be expected for a protein incapable or binding ATP due to a point mutation in the conserved lysine residue of its ATP-binding, Walker A motif (Figure 6.9A & Figure 6.10A). This data strongly argues against the possibility of an unidentified contaminant in the purified protein samples as being the source of the ATPase activity. Furthermore, the ATPase activity of SMARCAD1 is dependent on the presence of nucleosome (Figure 6.8A &B). Likewise, by itself, KAP1 is incapable of ATP hydrolysis, even when stimulated with nucleosomes. Taken collectively, the data strongly suggests that the nucleosome-stimulated ATPase activity is a specific activity of the SMARCAD1 enzyme.

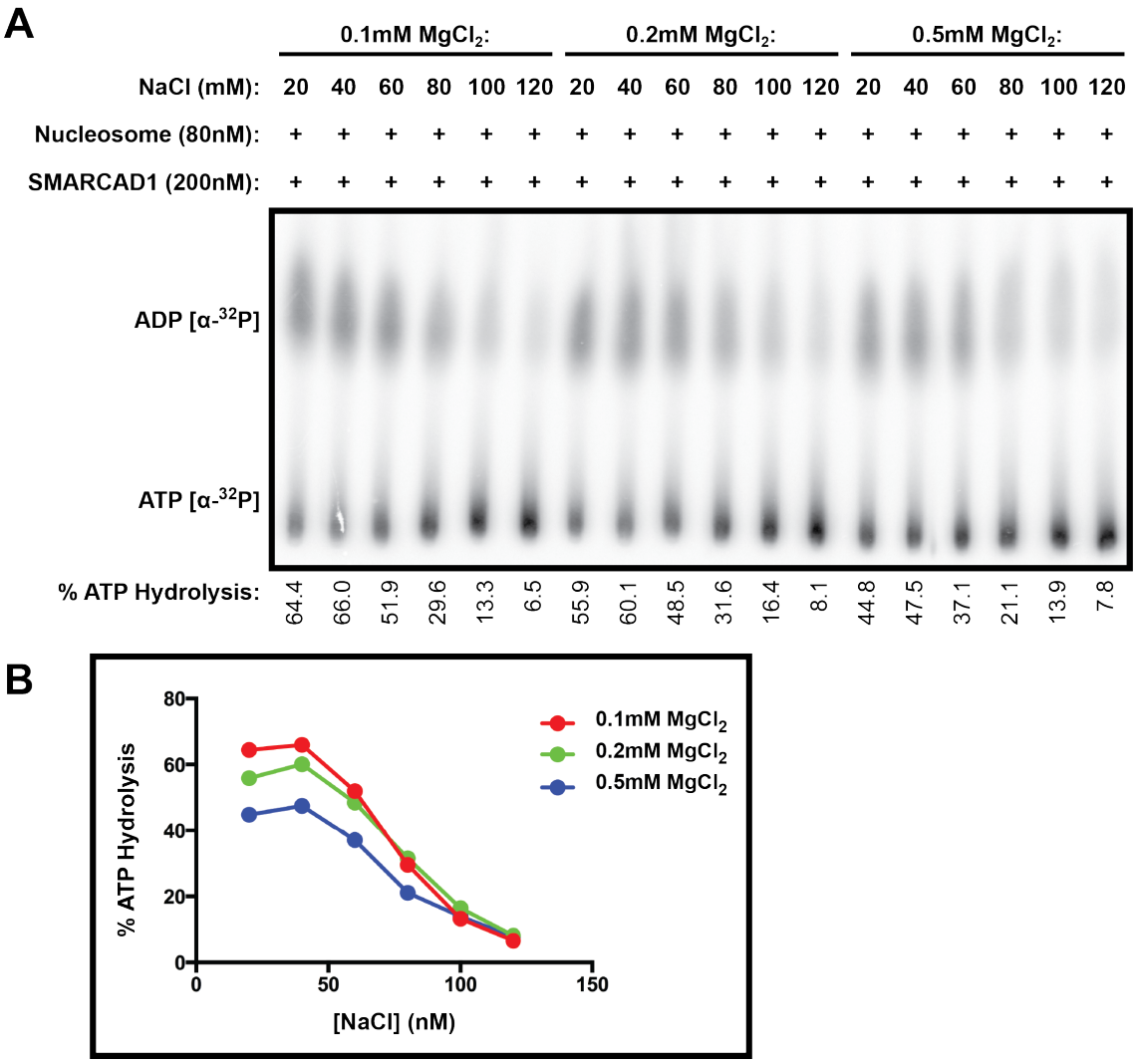


Figure 6.7 – Optimal Conditions for the ATPase Activity of SMARCAD1

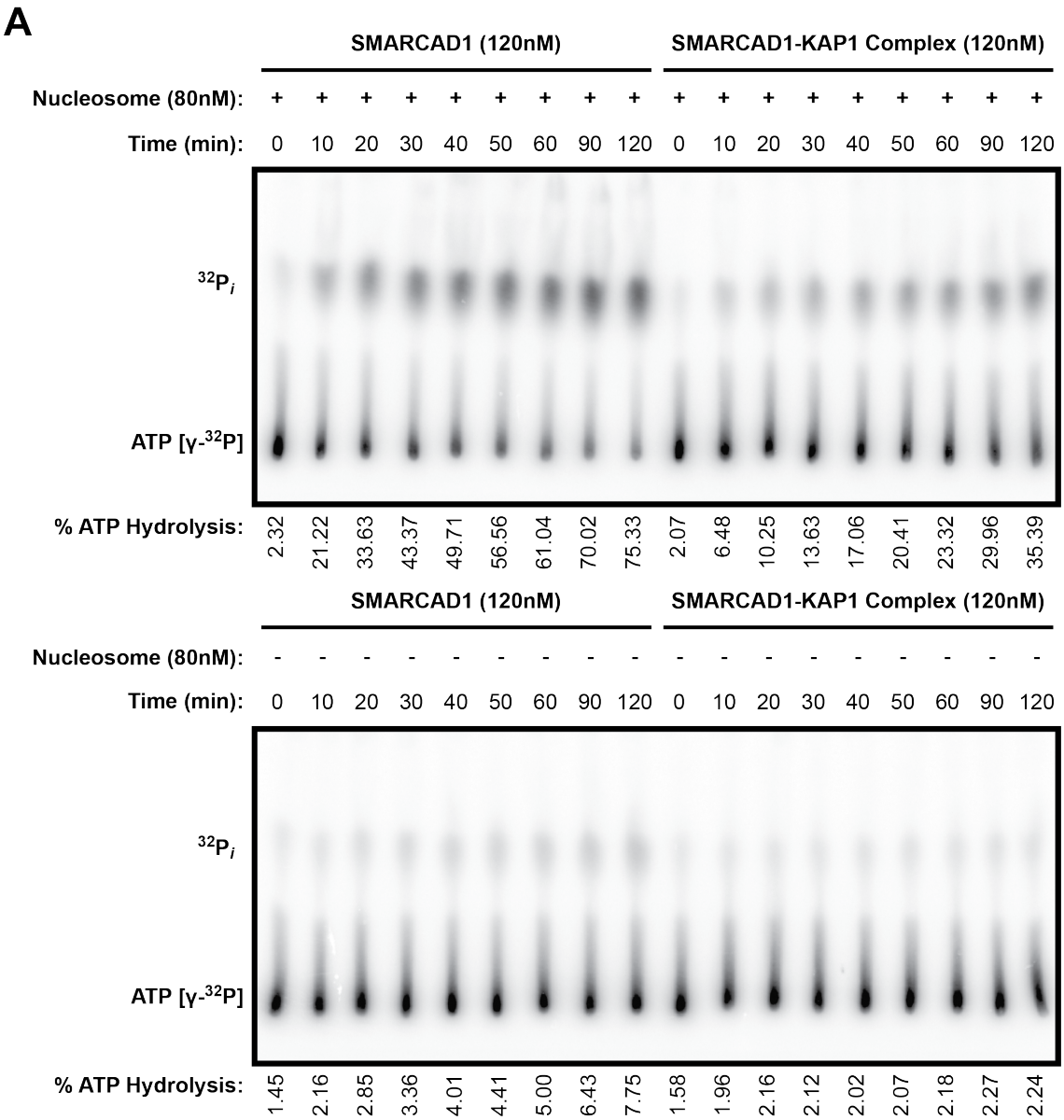
A. This series of ATPase reactions aims to determine the optimal concentrations of salt and magnesium for SMARCAD1 activity. Overall, it is evident that the nucleosome-stimulated ATPase activity of SMARCAD1 is exquisitely sensitive to the concentration of NaCl – increasing the NaCl concentration beyond an optimal of 40mM results in dramatically reduced enzymatic activity.

B. This graph plots the percentage of ATP hydrolysis observed for each condition. The red line relates to reactions containing 0.1mM $MgCl_2$, the green line indicates reactions with 0.2mM $MgCl_2$, while the blue line represents reactions with 0.5mM $MgCl_2$. The optimal concentration of NaCl is approximately 40mM; however, further increases in NaCl concentration result in a marked reduction in the ATPase activity of SMARCAD1. In contrast, the magnesium concentration does not have a major effect on the ATPase activity of SMARCAD1, though slightly more ATP hydrolysis is observed with 0.1mM $MgCl_2$ as compared to the other concentrations tested.

6.2.4 Comparison of the Nucleosome-Stimulated ATPase Activities of SMARCAD1 and the SMARCAD1-KAP1 Complex

To further validate the observation that KAP1 down-regulates the nucleosome-stimulated ATPase activity of SMARCAD1, the progress of the ATPase reaction with SMARCAD1 and the SMARCAD1-KAP1 was compared over time in a time course experiment (Figure 6.8). For both enzymes, there was increasing accumulation of the reaction product – in this case, hydrolysed inorganic phosphate – over time, though the amount of product appears to be reaching a plateau after two hours (Figure 6.8B). The amount of ATP hydrolysis observed with SMARCAD1 alone was higher than that for the SMARCAD1-KAP1 complex at all timepoints tested, indicating that the antagonism caused by KAP1 on the enzymatic activity of SMARCAD1 is reproducible.

The percentage of ATP hydrolysed under each condition was also quantified and plotted graphically to display the relationship between enzymatic activity and reaction duration for each of the enzymes (Figure 6.8B). The main distinction between the curves associated with SMARCAD1 and the SMARCAD1-KAP1 complex is the initial rate – the extent of ATP hydrolysis increases dramatically at the start of the reaction in the presence of SMARCAD1, whereas ATP hydrolysis by SMARCAD1-KAP1 proceeds much more sedately. In the latter stages of the reaction, however, the reaction rates appear largely comparable (Figure 6.8B). The slower reaction rate with increasing reaction duration, associated particularly with SMARCAD1 alone, may reflect either depletion of substrate – in this context, ATP – or enzymatic activity ‘dying’ over the course of the experiment, perhaps by protein denaturation. The former scenario may be possible since after two hours, the amount of un-hydrolysed ATP left in the reaction is approximately 25% of the original (Figure 6.8B). Nevertheless, Michaelis-Menten analysis, whereby the reaction rate is determined at different substrate concentrations, is necessary to distinguish between these two possibilities.


A
SMARCAD1 (120nM)
SMARCAD1-KAP1 Complex (120nM)
Nucleosome (80nM): + + + + + + + + + + + + + + + + +
Time (min): 0 10 20 30 40 50 60 90 120 0 10 20 30 40 50 60 90 120
32Pi
ATP [γ-32P]
% ATP Hydrolysis: 2.32 21.22 33.63 43.37 49.71 56.56 61.04 70.02 75.33 2.07 6.48 10.25 13.63 17.06 20.41 23.32 29.96 35.39
SMARCAD1 (120nM)
SMARCAD1-KAP1 Complex (120nM)
Nucleosome (80nM): - - - - - - - - - - - - - - - - -
Time (min): 0 10 20 30 40 50 60 90 120 0 10 20 30 40 50 60 90 120
32Pi
ATP [γ-32P]
% ATP Hydrolysis: 1.45 2.16 2.85 3.36 4.01 4.41 5.00 6.43 7.75 1.58 1.96 2.16 2.12 2.02 2.07 2.18 2.27 2.24
B
% ATP Hydrolysis
Time (min)
SMARCAD1 + Nucleosome
SMARCAD1 - Nucleosome
SMARCAD1-KAP1 + Nucleosome
SMARCAD1-KAP1 - Nucleosome

C

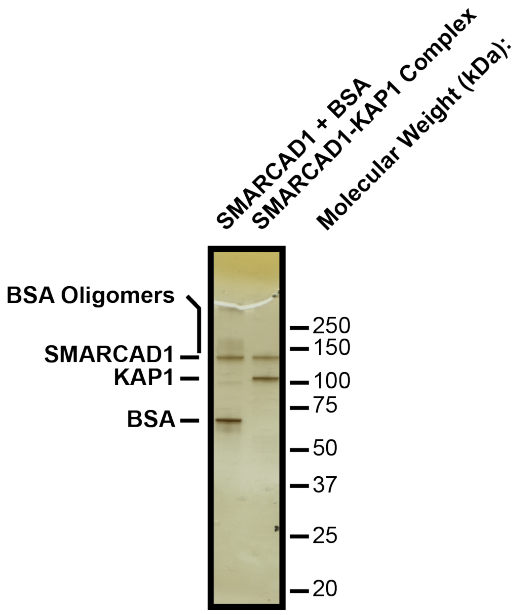


Figure 6.8 – Time Course Comparison of SMARCAD1 & the SMARCAD1-KAP1 Complex

A. The progress of the nucleosome-stimulated ATPase reaction by SMARCAD1 and the SMARCAD1-KAP1 complex was tracked over time here. Increasing levels of free hydrolysed γ-phosphate accumulated over time. At all timepoints, however, SMARCAD1 displayed higher levels of ATPase activity than the complex. As noted above, minimal ATP hydrolysis was observed in the absence of nucleosome (lower image).

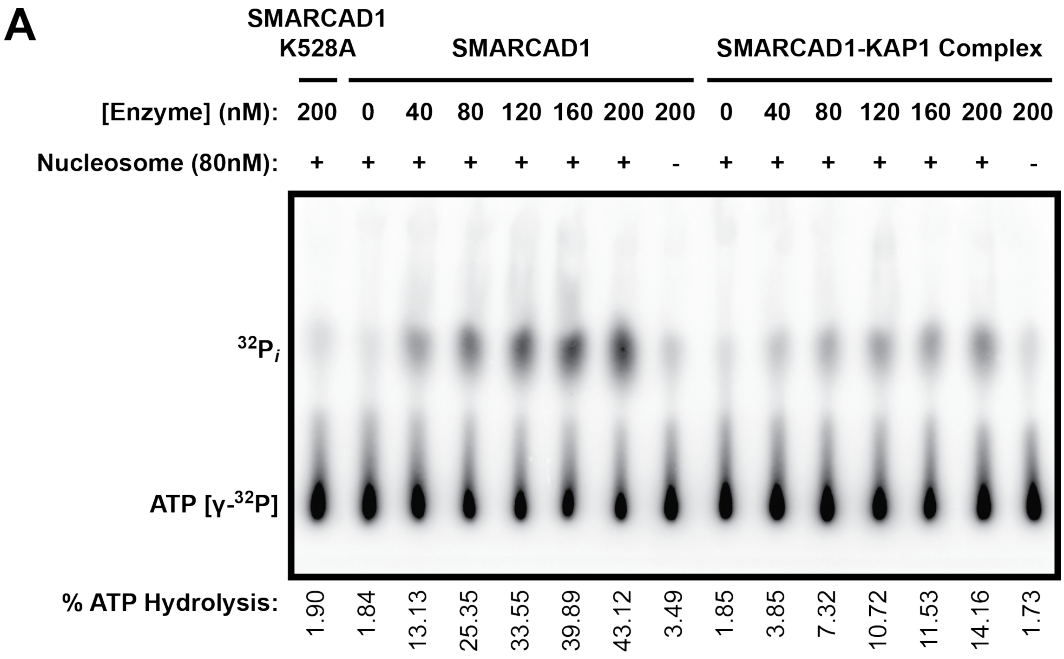
B. This graph plots the percentage of ATP hydrolysis observed at each timepoint for each condition. The red lines relate to reactions with SMARCAD1, while the blue lines relate to reactions with the SMARCAD1-KAP1 complex; likewise, the open diamonds reflect reactions performed in the absence of nucleosome, while the closed diamonds indicate the presence of nucleosome. Both reactions appear to be approaching a plateau by two hours. The two curves differ most prominently in the initial stages of the reaction as the initial reaction rate for SMARCAD1 is considerably higher than that for the SMARCAD1-KAP1 complex. In contrast, the rate at which hydrolysed ATP accumulates in the second-half of the reaction is comparable for both enzymes.

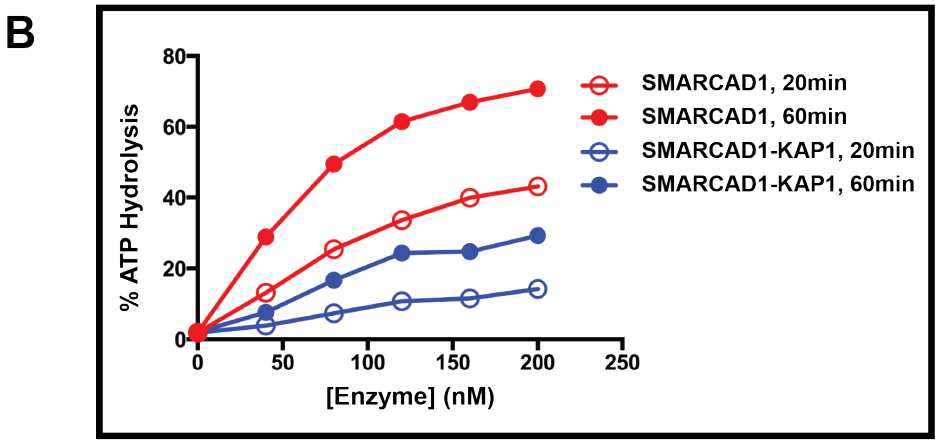
C. As assessed in this silver-stained gel, comparable concentrations of SMARCAD1 and the SMARCAD1-KAP1 complex were used for these reactions, suggesting that differences in activity cannot be attributed to differences in enzyme concentration.

Another implication of this time course experiment is that comparisons in enzymatic activity should be performed at a relatively early time point, as the considerable differences in initial rates between SMARCAD1 and the SMARCAD1-KAP1 complex would accentuate any differences in ATPase activity. Consequently, all further experiments featuring this assay focus upon the 20-minute timepoint. It is, nevertheless, important to emphasize that allowing these reactions to progress for 60 minutes did not materially alter the results or conclusions in any way (Figures 7.9B, 7.10B).

To further characterize the ATPase activities of SMARCAD1 and the SMARCAD1-KAP1 complex, a series of ATPase reactions were performed where increasing concentrations of enzyme were mixed with a constant concentration of recombinant nucleosome (Figure 6.9). For both SMARCAD1 and the SMARCAD1-KAP1 complex, as the enzyme concentration increased, the level of ATPase activity increased, though it appeared to be approaching a plateau at the highest enzyme concentrations tested (Figure 6.9B). Crucially, however, for any given enzyme concentration, the level of nucleosome-stimulated ATPase activity observed for the SMARCAD1-KAP1 complex was consistently lower than that of SMARCAD1 alone (Figure 6.9A & B). The inhibitory effect associated with KAP1 is not insignificant since the SMARCAD1-KAP1 complex generally exhibits three- to four-fold less nucleosome-stimulated ATPase activity compared to SMARCAD1 (Figure 6.9B).

Although the substrate in the ATPase assay is ATP, the ability of nucleosomes to stimulate SMARCAD1-mediated ATPase activity implies that SMARCAD1 acts on the nucleosome, which is a co-factor in the context of the ATPase assay. Thus, it might be expected that an increase in nucleosome concentration would result in a corresponding increase in ATP hydrolysis. Curiously, however, the relationship between nucleosome concentration and ATPase activities of SMARCAD1 and the SMARCAD1-KAP1 complex is biphasic (Figure 6.10A & B). At lower nucleosome concentrations, increases in cofactor concentration result in increased activity. As the nucleosome concentration approaches the enzyme concentration, however, further increases in cofactor concentration do not elicit additional increases in activity. Indeed, a large excess of nucleosome compared to the amount of SMARCAD1 or SMARCAD1-KAP1 complex may even be inhibitory.


A
SMARCAD1 K528A
SMARCAD1
SMARCAD1-KAP1 Complex
[Enzyme] (nM): 200 0 40 80 120 160 200 200 0 40 80 120 160 200 200
Nucleosome (80nM): + + + + + + + - + + + + + + -
32Pi
ATP [γ-32P]
% ATP Hydrolysis: 1.90 1.84 13.13 25.35 33.55 39.89 43.12 3.49 1.85 3.85 7.32 10.72 11.53 14.16 1.73


B
% ATP Hydrolysis
[Enzyme] (nM)
SMARCAD1, 20min
SMARCAD1, 60min
SMARCAD1-KAP1, 20min
SMARCAD1-KAP1, 60min

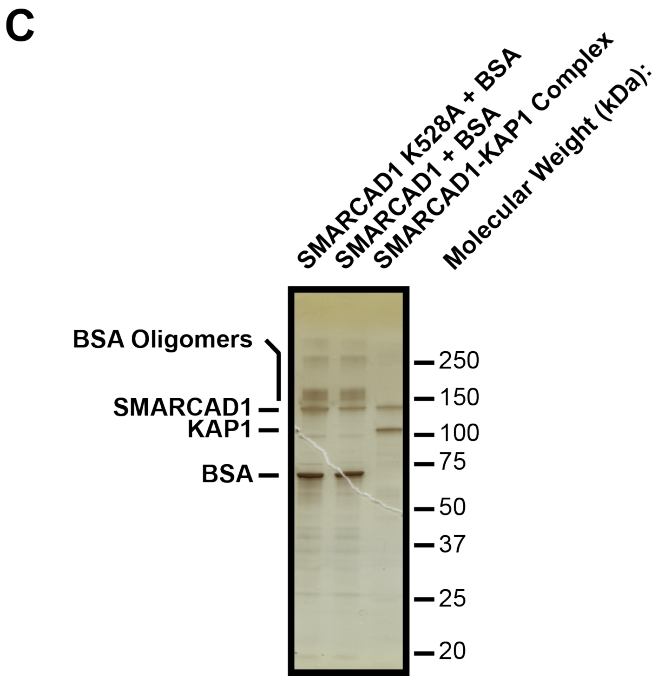

C
SMARCAD1 K528A + BSA
SMARCAD1 + BSA
SMARCAD1-KAP1 Complex
Molecular Weight (kDa):
BSA Oligomers
SMARCAD1
KAP1
BSA
250
150
100
75
50
37
25
20

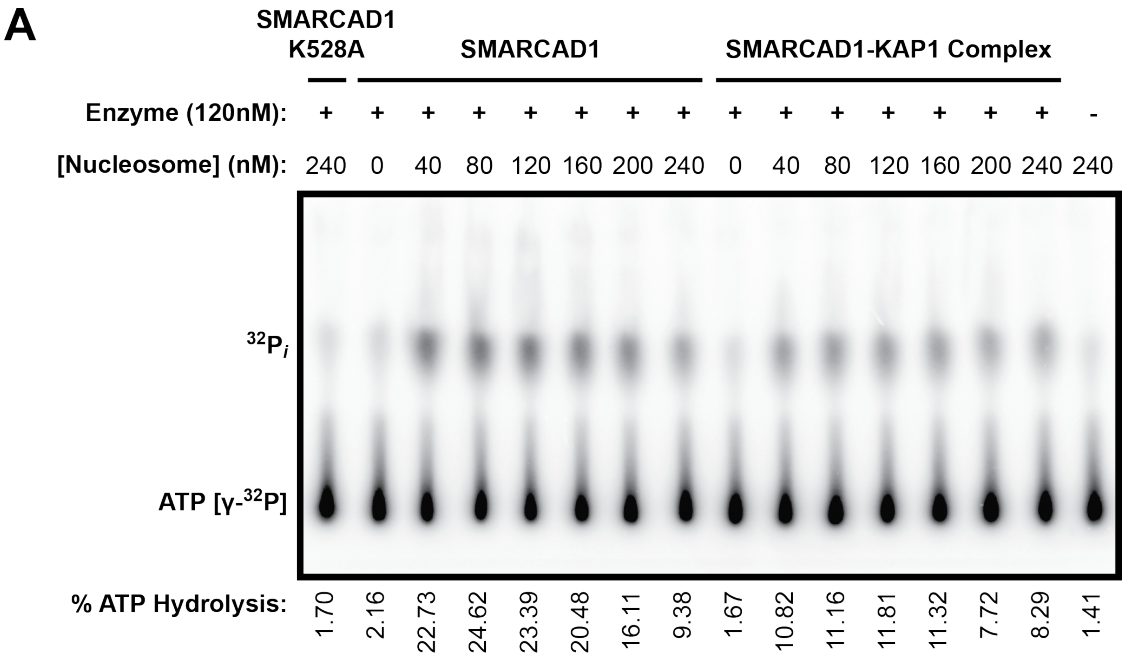
Figure 6.9 – Effect of Enzyme Concentration on the Nucleosome-Stimulated ATPase Activities of SMARCAD1 & the SMARCAD1-KAP1 Complex

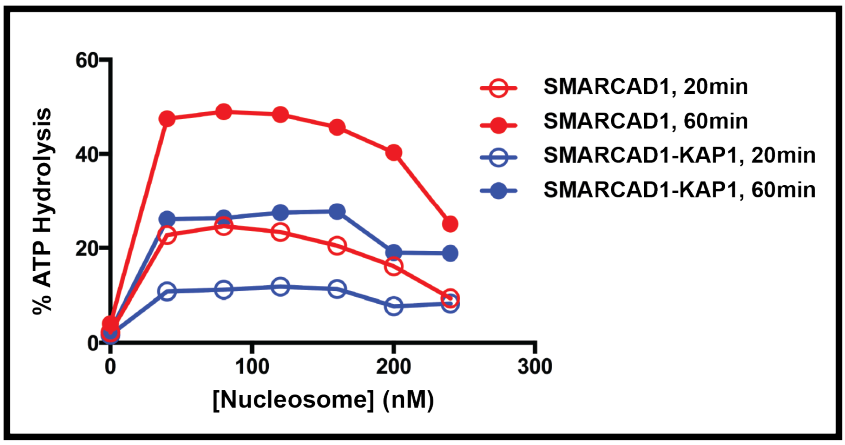
A. In this series of ATPase reactions, increasing concentrations of enzyme (either SMARCAD1 or the SMARCAD1-KAP1 complex) were combined with a constant amount of nucleosome (80nm). For both enzymes, increasing the concentration increased the amount of ATP hydrolysis observed. However, the SMARCAD1-KAP1 complex was consistently less active than SMARCAD1 alone at all concentrations tested. Moreover, the nucleosome-stimulated ATPase activity observed with these enzymes is specific as the 'ATPase-dead', K528A SMARCAD1 mutant displays no ATP hydrolysis even when stimulated with nucleosomes.

B. The percentage of ATP hydrolysis observed under each condition is plotted in this graph. The red lines relate to reactions with SMARCAD1, while the blue lines relate to reactions with the SMARCAD1-KAP1 complex; likewise, the open circles reflect 20-minute reactions, while the closed circles reflect 60-minute reactions. This graph shows that SMARCAD1 is more active than the reconstituted complex for any given enzyme concentration.

C. As assessed in this silver-stained gel, comparable concentrations of SMARCAD1, 'ATPase-dead' K528A SMARCAD1 mutant and the SMARCAD1-KAP1 complex were used for these reactions.

This behaviour is theoretically atypical since, by definition, enzymes should be able to catalyse reactions while remaining unaltered. However, one possible explanation is that the SMARCAD1 and SMARCAD1-KAP1 proteins gradually degrade over the course of the reaction; indeed, this possibility may be supported by the reduced reaction rates observed at late timepoints in a timecourse experiment as compared to that at early timepoints (Figure 6.8A & B). A second possibility is that the samples of purified SMARCAD1 and SMARCAD1-KAP1 complex contain only a small proportion of active enzyme. Thus, even the lower nucleosome concentrations tested would represent an excess to the amount of active enzyme present in the reactions. Yet, this explanation still fails to explain the inhibitory effect of a large excess of nucleosome. Moreover, a heparin column is a key step in the purification of SMARCAD1; as heparin is a DNA-mimetic, purification on the basis of affinity for heparin (and thus, indirectly, DNA) should be a reasonable means of enriching for active molecules. A third possibility is that the reconstituted nucleosome sample contains an inhibitory contaminant activity; at high nucleosome concentrations, the effect of this inhibitory contaminant outweighs that of the stimulatory activity, resulting in a net decrease in ATP hydrolysis.


A
SMARCAD1 K528A
SMARCAD1
SMARCAD1-KAP1 Complex
Enzyme (120nM): + + + + + + + + + + + + + + + -
[Nucleosome] (nM): 240 0 40 80 120 160 200 240 0 40 80 120 160 200 240 240
32Pi
ATP [γ-32P]
% ATP Hydrolysis: 1.70 2.16 22.73 24.62 23.39 20.48 16.11 9.38 1.67 10.82 11.16 11.81 11.32 7.72 8.29 1.41


B
% ATP Hydrolysis
60
40
20
0
0
100
200
300
[Nucleosome] (nM)
SMARCAD1, 20min
SMARCAD1, 60min
SMARCAD1-KAP1, 20min
SMARCAD1-KAP1, 60min

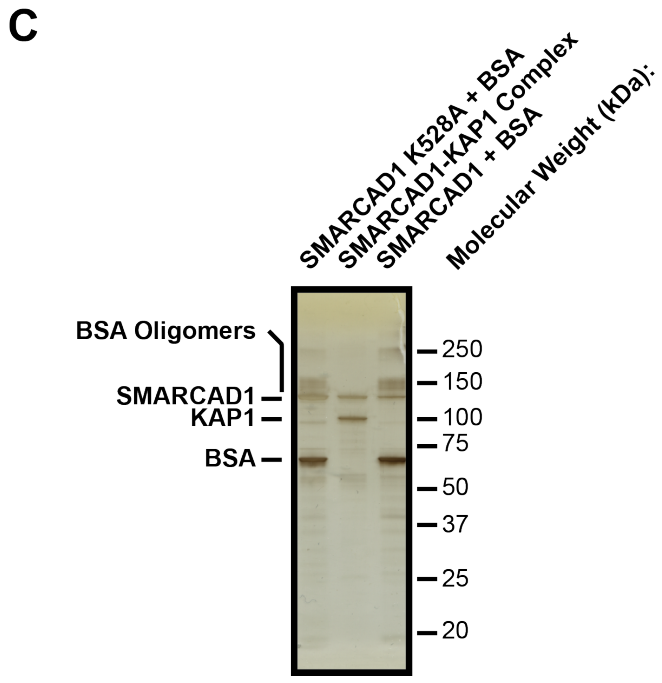

C
SMARCAD1 K528A + BSA
SMARCAD1-KAP1 Complex
SMARCAD1 + BSA
Molecular Weight (kDa):
BSA Oligomers
SMARCAD1
KAP1
BSA
250
150
100
75
50
37
25
20

Figure 6.10 – Effect of Nucleosome Concentration on the ATPase Activities of SMARCAD1 & the SMARCAD1-KAP1 Complex

A. In this series of ATPase reactions, increasing concentrations of recombinant nucleosomes were combined with a constant concentration of enzyme (120nM). Virtually no ATPase activity is observed when SMARCAD1 is not stimulated with nucleosome. At low nucleosome concentrations, further increases result in greater activity. However, as the nucleosome concentration approximates the SMARCAD1 concentration, further increases are not accompanied by additional increases in activity; indeed, an excess of nucleosome may even be inhibitory. Nonetheless, SMARCAD1 is consistently more active than the SMARCAD1-KAP1 complex.

B. The percentage of ATP hydrolysis observed under each condition is plotted in this graph. The red lines relate to reactions with SMARCAD1, while the blue lines relate to reactions with the SMARCAD1-KAP1 complex; likewise, the open circles reflect 20-minute reactions, while the closed circles reflect 60-minute reactions. This graph emphasizes the biphasic effect of nucleosome concentration on the ATPase activities of SMARCAD1 and the SMARCAD1-KAP1 complex. At low nucleosome concentrations, further increases in substrate result in more activity; above a point, however, additional nucleosomes in the reaction may even be inhibitory.

C. As assessed in this silver-stained gel, comparable concentrations of SMARCAD1, the 'ATPase-dead' K528A SMARCAD1 mutant, and the SMARCAD1-KAP1 complex were used for these reactions.

In spite of the curious relationship between nucleosome concentration and ATPase activities of SMARCAD1 and the SMARCAD1-KAP1 complex, all the data presented here are, nonetheless, unanimous in showing that KAP1 acts to inhibit or down-regulate the enzymatic activity of SMARCAD1.

Finally, it was observed that KAP1 added directly to an ATPase reaction with SMARCAD1 did not exert an inhibitory effect. This is likely to be an experimental artefact reflecting the relatively slow rate of complex formation; indeed, we note that the efficiency of reconstituting the SMARCAD1-KAP1 complex is low, unless long incubation periods are allowed for the affinity purification steps.

6.3 Conclusions

In this part of the study, we have described protocols to purify recombinant SMARCAD1 and KAP1 expressed in *E. coli* to a high level of purity. Moreover, it has been possible to reconstitute the SMARCAD1-KAP1 complex by way of sequential affinity purification steps. Importantly, the reconstituted SMARCAD1-KAP1 complex behaves as a *bona fide* protein complex – when analysed by gel filtration chromatography, SMARCAD1 and KAP1 precisely co-elute with each

other. Furthermore, the complex elutes earlier in the size exclusion gradient than each individual component, as would be expected for a larger molecule.

We have also shown, for the first time, that the SMARCAD1 protein is capable of ATP hydrolysis, particularly in the presence of nucleosomes. Fittingly, a substrate preference for nucleosomes is characteristic of many chromatin remodellers. While the SMARCAD1-KAP1 complex is also optimally stimulated by nucleosomes, we show in several experiments that it is consistently less active at ATP hydrolysis than SMARCAD1 by itself. The inhibitory effect of KAP1 is not insignificant – in general, it results in a three- to four-fold reduction in nucleosome-stimulated ATP hydrolysis. Yet, this data cannot definitively indicate the mechanism by which KAP1 antagonises the ATPase activity of SMARCAD1. It is, however, likely that the reduction in initial reaction rate observed in the presence of KAP1 significantly contributes to this antagonism. Nevertheless, since ATP hydrolysis is a surrogate measure for a complex, probably multi-step, enzymatic reaction, it is impossible to distinguish between a reduced affinity for nucleosomes or a reduction in processivity amongst other possibilities. It would be extremely useful for the preliminary characterization of the SMARCAD1 and SMARCAD1-KAP1 enzymes were to be extended by Michaelis-Menten analysis; for example, such analysis should help distinguish between depletion of substrate and protein degradation as the reason for the plateauing ATP hydrolysis observed at late timepoints. Also, the use of a malachite green ATPase assay, where phosphate release is measured, might be more suitable in subsequent experiments since it allows ATP hydrolysis to be quantified more accurately and is amenable to being performed in high-throughput formats. Further experiments such as electromobility shift assays to determine the dissociation constant between SMARCAD1 or the SMARCAD1-KAP1 complex and nucleosome, or nucleosome sliding assays will be required to further elucidate upon the mechanism by which KAP1 acts as an inhibitory factor.

Chapter 7. Results IV: Characterization of the Interaction between SMARCAD1 & KAP1

7.1 Aims

In addition to the conserved helicase-like ATPase domain, chromatin remodellers often possess other accessory domains that either enhance or confer specificity to the basal enzymatic activity of the protein. These domains are generally, though not universally, histone-mark readers. With regards to the SMARCAD1 protein, however, the only other annotated domains are a pair of tandem CUE domains in its C-terminus. As CUE domains are ubiquitin-binding domains, the rationale for their presence in a chromatin remodeller is not immediately apparent. Thus, an important question that we wanted to address was the role of the SMARCAD1 CUE domains.

Although SMARCAD1 and KAP1 exist constitutively as a stable complex, based on their annotated protein domains, there is no *prima facie* indication for the means by which this interaction is mediated. However, as previously discussed, we were able to reconstitute the interaction with purified, recombinant proteins. To gain further insight into the nature of the SMARCAD1-KAP1 interaction and the means by which it could be regulated, we wanted to biochemically characterize this interaction and to define the regions of the two proteins that mediate it.

7.2 Results

7.2.1 Mutation of the CUE Domains of SMARCAD1 Abrogates the SMARCAD1-KAP1 Interaction in Cells

Our initial approach to investigate the function of the CUE domains of SMARCAD1 involved a T-Rex™ 293 cell line that was depleted of endogenous SMARCAD by shRNA knockdown, but inducibly expressed exogenous SMARCAD1 under the control of a doxycycline regulated promoter. (This cell line was generated by Hannah Williams.) By titration of the doxycycline concentration, SMARCAD1 was expressed to approximately endogenous levels (Figure 7.1A). A comparable cell

line rescued with exogenous SMARCAD1 bearing four point mutations in each of its tandem CUE domains was also established. As these alanine substitutions target the MFP and LL motifs, this SMARCAD1 mutant should be rendered entirely incapable of binding to ubiquitin.

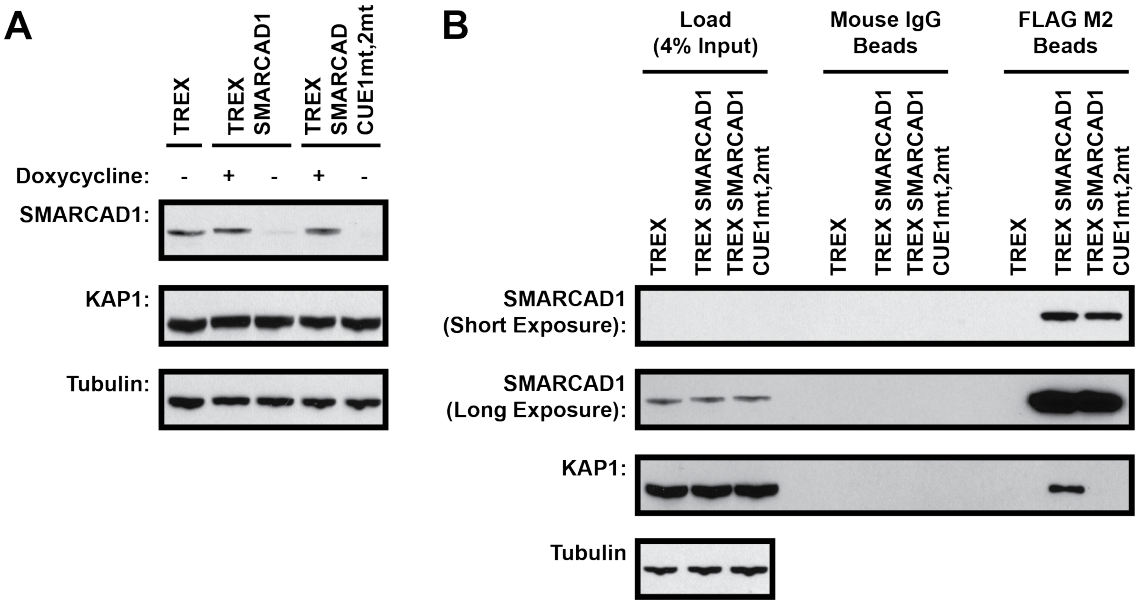


Figure 7.1 – Mutation of the CUE Domains of SMARCAD1 Abrogates the SMARCAD1-KAP1 Interaction in Cells

A. Endogenous SMARCAD1 was extensively depleted (see '-' doxycycline samples) from parental T-REx™ 293 cells (Invitrogen) by shRNA. These cells were then complemented with either wild-type (T-REx SMARCAD1) or CUE domain mutant SMARCAD1 (T-REx SMARCAD1 CUE1mt,2mt). As a doxycycline-regulated promoter controlled expression of exogenous SMARCAD1, it was possible to rescue these cells with approximately endogenous levels of SMARCAD1. The CUE domain mutant possesses four point mutations in each of its two CUE domains, which consequently, should be functionally inactive.

B. Relatively equal levels of wild-type SMARCAD1 or the double CUE domain mutant were pulled down by immunoprecipitation with FLAG M2 agarose from the cell lines described above. While KAP1 co-immunoprecipitated with SMARCAD1, this interaction was completely abrogated by mutation of the tandem CUE domains of SMARCAD1.

To determine whether mutation of the CUE domains of SMARCAD1 affects its interaction partners, both wild-type SMARCAD1 and the double CUE domain mutant were pulled down from the aforementioned cell lines by immunoprecipitation with M2 agarose. As expected, wild type SMARCAD1 co-immunoprecipitated with KAP1. Notably, this interaction was abrogated in the CUE domain mutant (Figure 7.1B). (This observation was initially made by Hannah Williams, but the data presented here was reproduced in my hands.) Importantly, both cell lines expressed comparable levels of KAP1, indicating that the inability of

the SMARCAD1 CUE domain mutant to co-immunoprecipitate with KAP1 was not because of dramatic differences in KAP1 expression.

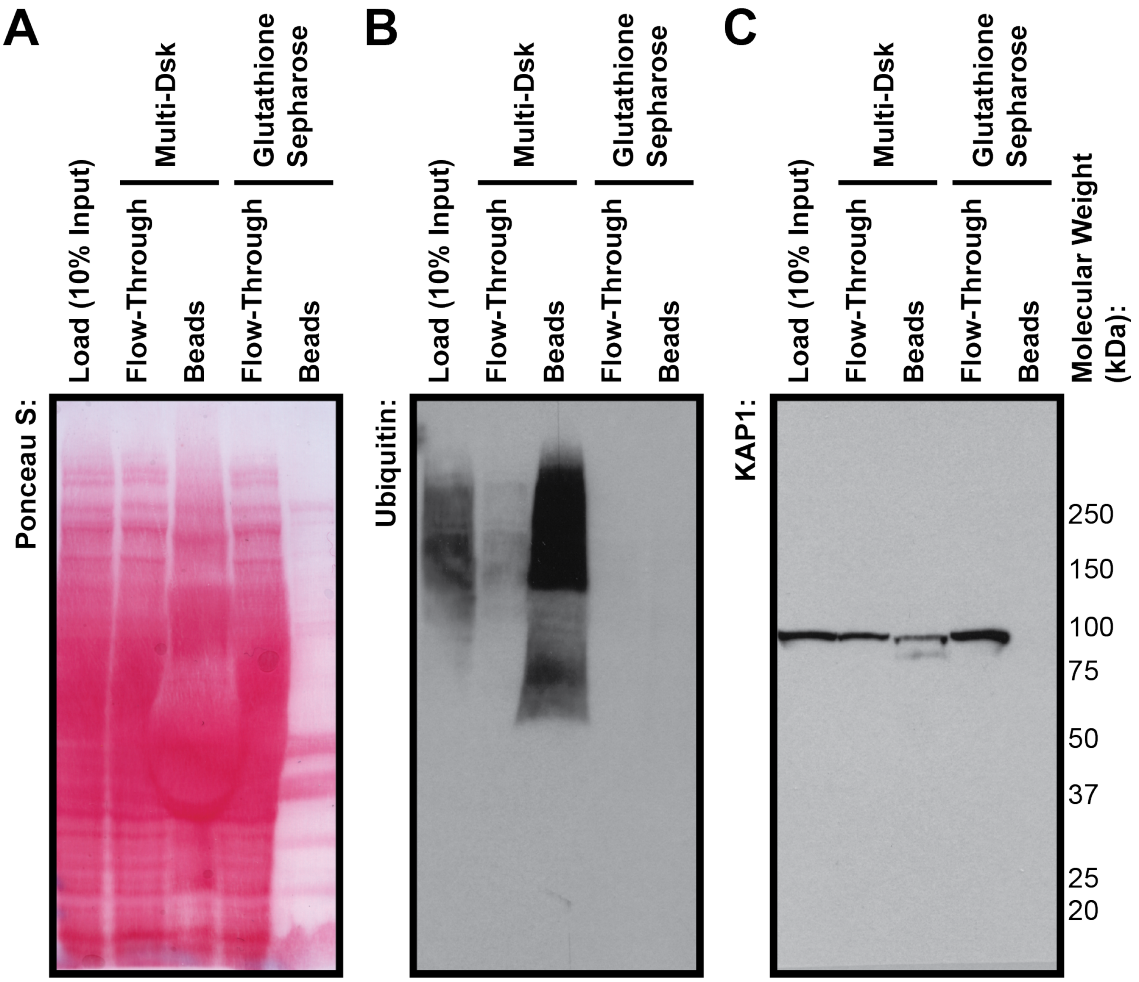


Figure 7.2 – KAP1 is Not Constitutively Ubiquitylated in Human Cells

A. Multi-Dsk, an affinity resin with a high avidity for ubiquitin, was used to enrich for ubiquitylated proteins from 293T cell extract. The enrichment for ubiquitylated proteins by the Multi-Dsk resin over non-specific interactions with the control glutathione sepharose beads can be visualized by Ponceau S staining.

B. A Western blot against ubiquitin also demonstrates the specific enrichment for ubiquitylated proteins by the Multi-Dsk resin.

C. There is some KAP1 binding to the Multi-Dsk resin; however, the absence of higher molecular weight forms of KAP1 suggests that KAP1 is not constitutively ubiquitylated in human 293T cells.

As SMARCAD1 interacts constitutively with KAP1, and as CUE domains are ubiquitin-binding domains, it was hypothesized that KAP1 would be constitutively ubiquitylated. Taking advantage of the high avidity for ubiquitin of the Multi-Dsk resin, a fusion protein consisting of five repeats of the UBA ubiquitin-binding domain of Dsk2 (Wilson et al., 2012), ubiquitylated proteins were enriched from

293T human cell extract (Figure 7.2A & B). By Western blotting, it was identified that some KAP1 bound to the Multi-Dsk resin, though there were no KAP1 bands of a higher molecular weight, as would be expected for ubiquitylation (Figure 7.2C). This data strongly argues against the existence of constitutive KAP1 ubiquitylation. The apparent discrepancy of KAP1 binding to the Multi-Dsk resin in the absence of ubiquitin might reflect, for example, the indirect binding via an intermediary, ubiquitylated protein; this explanation would be consistent with the mild stringency of this immunoprecipitation experiment.

7.2.2 SMARCAD1-KAP1 Interaction Depends on the First CUE Domain of SMARCAD1 and the RBCC Domain of KAP1 *In Vitro*

As described above, the SMARCAD1-KAP1 interaction can be reconstituted *in vitro* with purified, recombinant proteins (Figure 6.4). Thus, to further investigate whether the reconstituted interaction truly recapitulates that observed in cells, three additional SMARCAD1 mutants were purified to assess whether the same dependence on the CUE domains for the interaction exists *in vitro*. These N-terminally FLAG-tagged SMARCAD1 mutants had point mutations in individual or both CUE domains and consequently, have compromised ubiquitin-binding ability in one or both of its CUE domains (Figure 7.3A). The ability of the SMARCAD1 CUE domain mutants to bind to KAP1 *in vitro* was compared to that of wild type SMARCAD1 in a binding assay (Figure 7.3B). Following overnight incubation of the recombinant proteins, pulling down wild type SMARCAD1 resulted in co-immunoprecipitation of stoichiometric levels of KAP1. In contrast, however, the mutant with point mutations in the first CUE domain was severely compromised in its ability to bind KAP1. Predictably, the double mutant behaved similarly, thereby recapitulating the phenotype observed in cells. Nevertheless, mutation of the second CUE domain did not perceptibly affect the ability of SMARCAD1 to bind KAP1. Hence, this data indicates that the SMARCAD1-KAP1 interaction is strictly mediated via the first CUE domain of SMARCAD1 and that the tandem CUE domains of SMARCAD1 are not functionally redundant (Figure 7.3B).

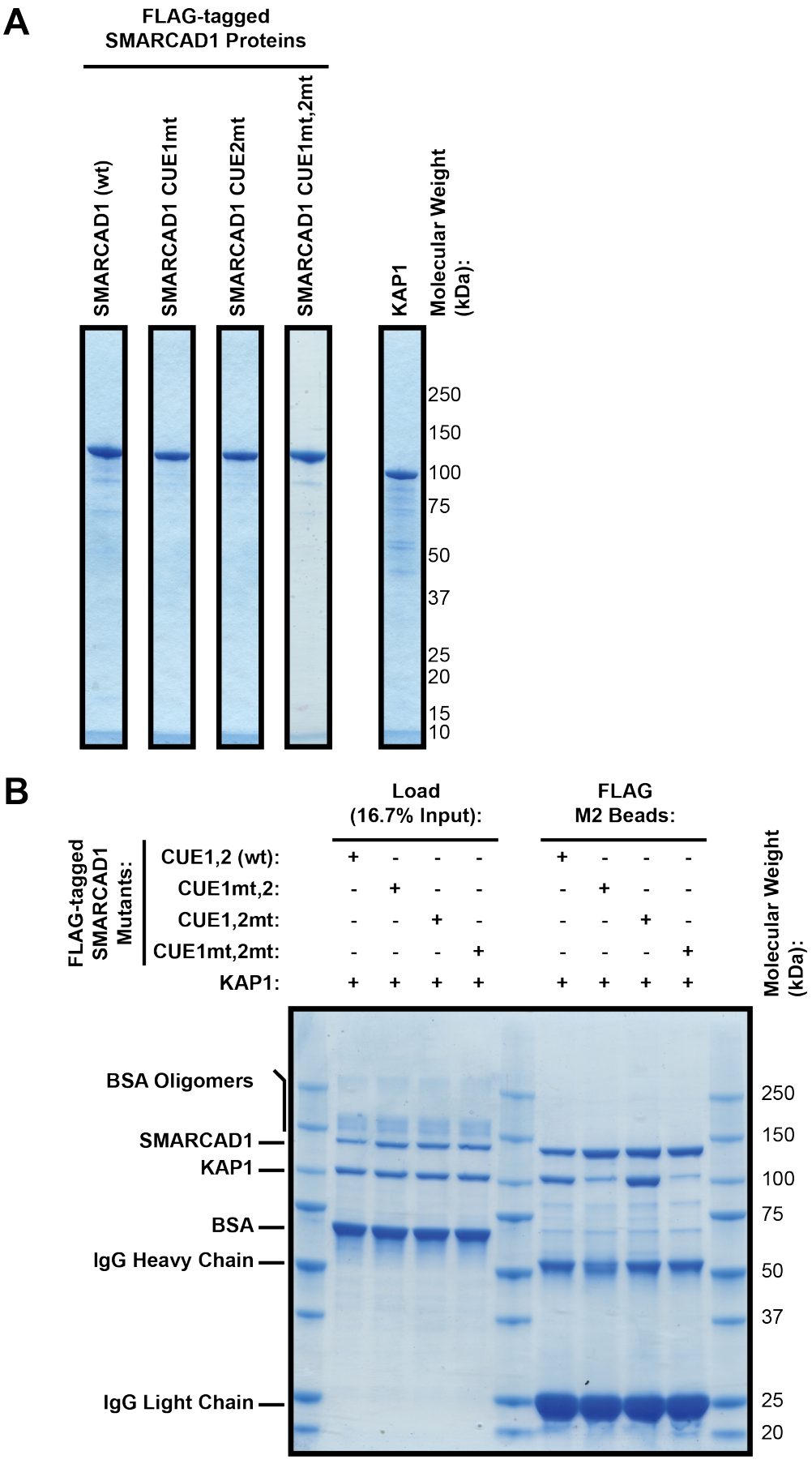

A
FLAG-tagged SMARCAD1 Proteins
SMARCAD1 (wt)
SMARCAD1 CUE1mt
SMARCAD1 CUE2mt
SMARCAD1 CUE1mt,2mt
KAP1
Molecular Weight (kDa):
250
150
100
75
50
37
25
20
15
10
B
Load (16.7% Input):
FLAG M2 Beads:
FLAG-tagged SMARCAD1 Mutants:
CUE1,2 (wt):
CUE1mt,2:
CUE1,2mt:
CUE1mt,2mt:
KAP1:
Molecular Weight (kDa):
BSA Oligomers
SMARCAD1
KAP1
BSA
IgG Heavy Chain
IgG Light Chain
250
150
100
75
50
37
25
20

Figure 7.3 – The SMARCAD1-KAP1 Interaction Depends on the First CUE Domain of SMARCAD *In Vitro*

A. As assessed by InstantBlue staining, the purified, recombinant protein samples used to reconstitute the SMARCAD1-KAP1 interaction are of a high purity. In addition to wild type full-length SMARCAD1, three additional mutants, which had point mutations in individual or both CUE domains, were purified.

B. Purified, recombinant wild type and mutant FLAG-tagged SMARCAD1 proteins were mixed with purified, recombinant KAP1 along with BSA. The different versions of SMARCAD1 were immunoprecipitated by FLAG M2 agarose to determine whether or not they possessed the ability to bind to KAP1. The SMARCAD1-KAP1 interaction was once again observed. Mutation of the first CUE domain severely compromised the ability of SMARCAD1 to bind to KAP1, while mutation of the second CUE domain did not perturb binding.

To determine the region of KAP1 responsible for binding to SMARCAD1, truncated fragments of KAP1 were assessed for retention of the ability to specifically interact with the first CUE domain of SMARCAD1. The initial KAP1 fragments truncated the protein along annotated protein domains; however, they did not offer much insight apart from demonstrating that the C-terminal PHD-bromodomain was dispensable for the interaction with SMARCAD1. Thus, limited tryptic digestion was used to identify KAP1 fragments that reflect the actual structure of the protein. By incubating SMARCAD1 with a limiting concentration of trypsin for a relatively short period of time, three fragments of KAP1 that displayed a comparatively higher resistance to trypsin were identified (Figure 7.4A). These trypsin-resistant KAP1 fragments probably represent ordered regions of the protein and consequently, are more likely to contain the interaction interface with SMARCAD1. To map these fragments, the sequences of their N-termini were determined by Edman degradation, while their mass were measured to a high level of accuracy by intact molecular weight mass spectrometry. Interpreted in line with the knowledge that trypsin cleaves primarily after lysine or arginine residues, the largest KAP1 fragment (S33-K434, 45kDa) was identified to span the RBCC domain of KAP1 (Figure 7.4B). Further cleavage yields the second fragment, which encompasses the second B-box and the coiled-coil domain (S200/D202-K434, 27kDa). The final fragment (L592-P835) covers the C-terminal PHD-bromodomain.

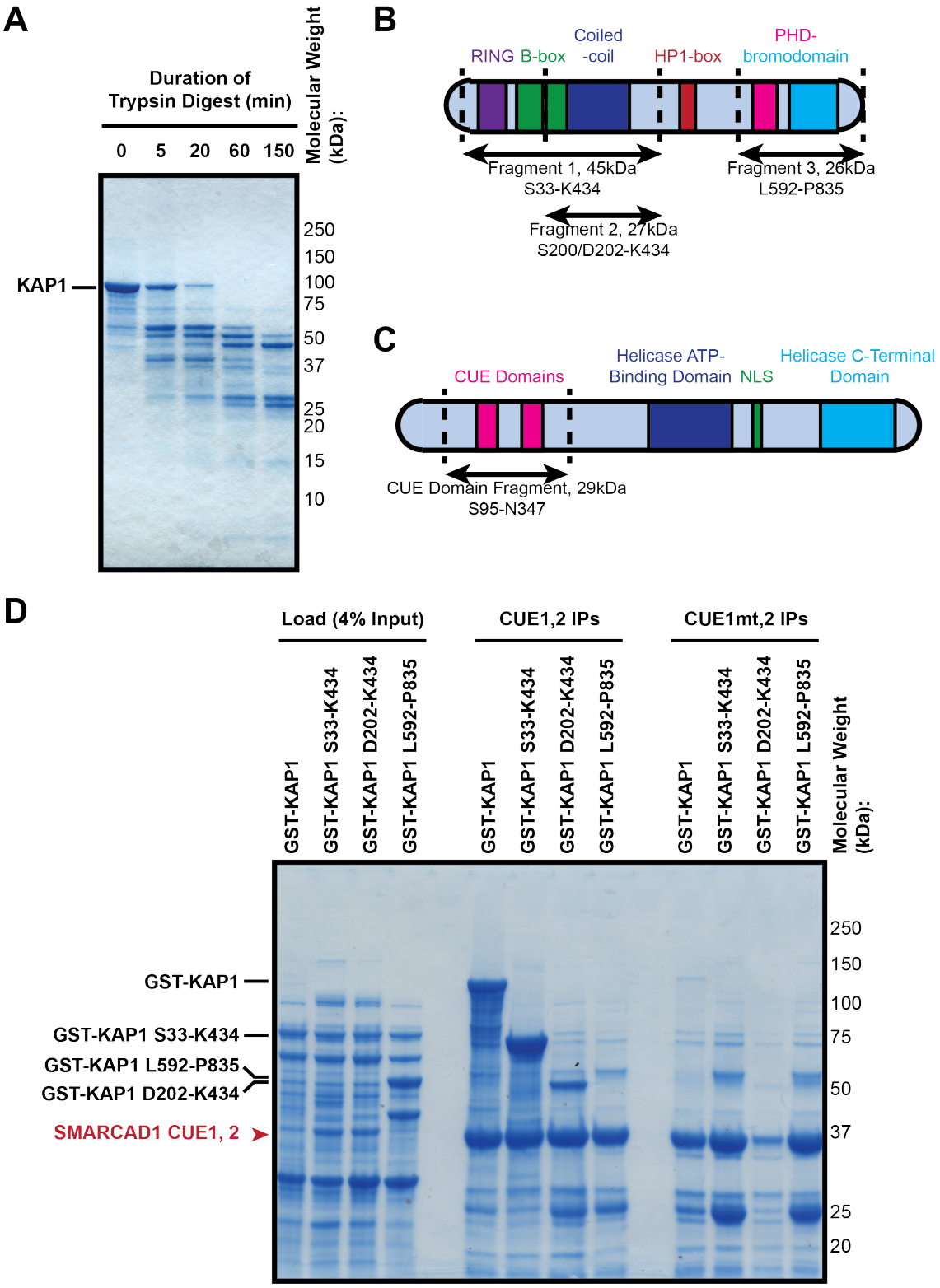


Figure 7.4 – The Trypsin-Resistant KAP1 S33-K434 Fragment Binds Specifically to the First CUE Domain of SMARCAD1

A. Pure recombinant KAP1 was incubated with a limiting concentration of trypsin to identify fragments that were resistant to tryptic digestion. Three relatively stable fragments are obtained after 2.5 hours of digestion at 20°C. These fragments probably represent ordered regions of the KAP1 protein.

B. The three trypsin-resistant KAP1 fragments were mapped by a combination of Edman degradation and intact molecular weight mass spectrometry. These complementary techniques reveal the sequences of the N-termini of the fragments and their precise mass respectively. The largest fragment, S33-K434, spans the entire RBCC domain of KAP1. It can be further digested into a small fragment, S200/D202-K434, which encompasses the second B-box and the coiled-coil domain. The C-terminal PHD-bromodomain forms the final fragment, L592-P835.
C. The SMARCAD1 CUE domain fragment, S95-N347, is illustrated in this schematic of the SMARCAD1 protein.
D. Together with full-length KAP1, the three trypsin-resistant fragments were expressed in *E. coli* as GST fusion proteins. Bacterial protein extracts containing these GST fusion proteins were then incubated with beads immobilised with a CUE domain fragment of SMARCAD1. The CUE domain SMARCAD1 fragment strongly enriched for full-length KAP1 and the S33-K434 RBCC fragment. This interaction was specific as the control CUE1mt,2 fragment, which contains a mutated first CUE domain, failed to bind either full-length protein or KAP1 S33-K434. Some enrichment of D202-K434 was also observed, slbeit to a lesser extent.

Along with full-length KAP1, the three trypsin-resistant fragments were expressed in *E. coli* as GST fusion proteins. Bacterial protein extracts containing these GST fusion proteins were then incubated with beads immobilised with a CUE domain fragment of SMARCAD1 (Figure 7.4C). The CUE domain SMARCAD1 fragment strongly enriched for full-length KAP1 and the S33-K434 RBCC fragment (Figure 7.4D). Crucially, this interaction depended on the integrity of the first CUE domain of SMARCAD1; mutation completely abrogated binding. Enrichment to a lesser extent was also observed for the D202-K434 fragment; the apparent reduction in affinity may, in fact, simply reflect lower abundance in extract due to significantly poorer solubility. Strikingly, although the RBCC domain is present in its entirety, the M1-E388 fragment does not bind to the SMARCAD1 CUE domain fragment, indicating that the forty-seven amino acids immediately after the RBCC domain are required for KAP1 to retain the ability to interact with SMARCAD1. It is, however, unclear whether those forty-seven residues directly contact the SMARCAD1 CUE domain, or contribute to proper folding of the protein.

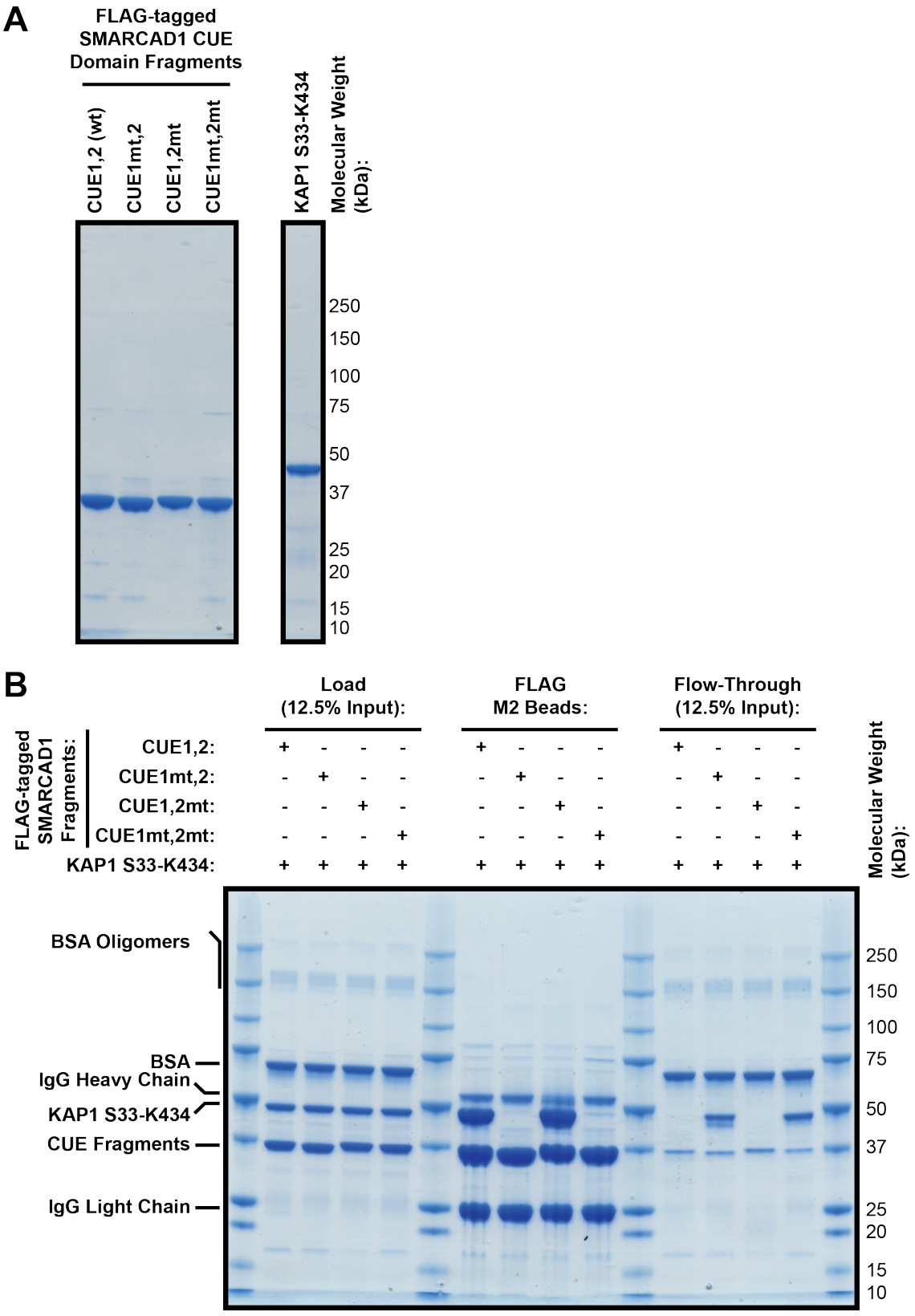


Figure 7.5 – Reconstitution of the Minimal SMARCAD1-KAP1 Interaction

A. InstantBlue staining was used to assess the purity of purified, recombinant wild type and mutant SMARCAD1 CUE domain fragments. Pure KAP1 S33-K434 is also shown in this figure.

B. Purified KAP1 S33-K434 binds specifically to the SMARCAD1 CUE domain fragment but not fragments with mutations in the first CUE domain. This suggests that the interaction between the minimal soluble fragment of KAP1, S33-K434, and the CUE domain fragment of SMARCAD1 is representative of the SMARCAD1-KAP1 interaction.

To further confirm that KAP1 S33-K434 is responsible for the interaction with SMARCAD1, we sought to repeat it with purified components. The GST-KAP1 S33-K434 fusion protein suffers from poor solubility and cannot be purified. However, its solubility is enhanced considerably when expressed using the pET28a-SUMO expression system, and can be purified by a combination of Ni-NTA affinity chromatography and anion exchange chromatography (Figure 7.5A). It is worth noting that despite several rounds of optimization, the D202-K434 fragment remains largely insoluble and intractable to purification. Hence, we regard the S33-K434 fragment as the minimal soluble fragment of KAP1 capable of interacting with SMARCAD1. In addition to the aforementioned SMARCAD1 CUE domain fragment (Figure 7.4C), three additional mutants – with mutations in individual or both CUE domains – were purified (Figure 7.5A).

The KAP1 S33-K434 fragment was incubated briefly with the wild type SMARCAD1 CUE domain fragment or the various mutants before the CUE domain fragments were immunoprecipitated with FLAG M2 agarose. Stoichiometric quantities of KAP1 S33-K434 bound to the wild-type CUE domain fragment. Moreover, the KAP1 S33-K434 was completely incapable of interacting with either the CUE1mt,2 mutant, which has point mutations in the first CUE domain, or the CUE1mt,2mt double mutant. KAP1 S33-K434 bound just as efficiently to the CUE1,2mt mutant that possesses a mutated second CUE domain, as to the wild type CUE fragment, recapitulating the strict requirement for the first CUE domain of SMARCAD1 observed with full-length proteins (Figure 7.5B). Considered holistically, the data strongly suggests that the SMARCAD1-KAP1 interaction occurs at the level of a direct protein-protein interaction between the first CUE domain of SMARCAD1 directly and the N-terminal, S33-K434 portion of KAP1 encompassing the RBCC domain. The only caveat, however, is that the interaction observed between the protein fragments occurs with faster kinetics than that between full-length SMARCAD1 and KAP1 – strong enrichment of the KAP1 S33-K434 fragment by the SMARCAD1 CUE domain was observed after incubation for an hour, while

efficient reconstitution of the SMARCAD1-KAP1 complex was only seen after overnight incubation.

The poor solubility of KAP1 fragments involving the RBCC domain (e.g. the D202-K434 fragment which appears to retain some ability to bind to SMARCAD1) imposes significant limitations on any attempt to more precisely map the interaction interface used by KAP1 to bind SMARCAD1 by truncation analysis. Consequently, a peptide array spanning the KAP1 sequence was used to determine whether the SMARCAD1 CUE domain fragment recognises a specific KAP1 peptide. Since the array is composed of peptides twenty residues in length that progress along the KAP1 sequence in increments of a single amino acid, only linear epitopes are represented.

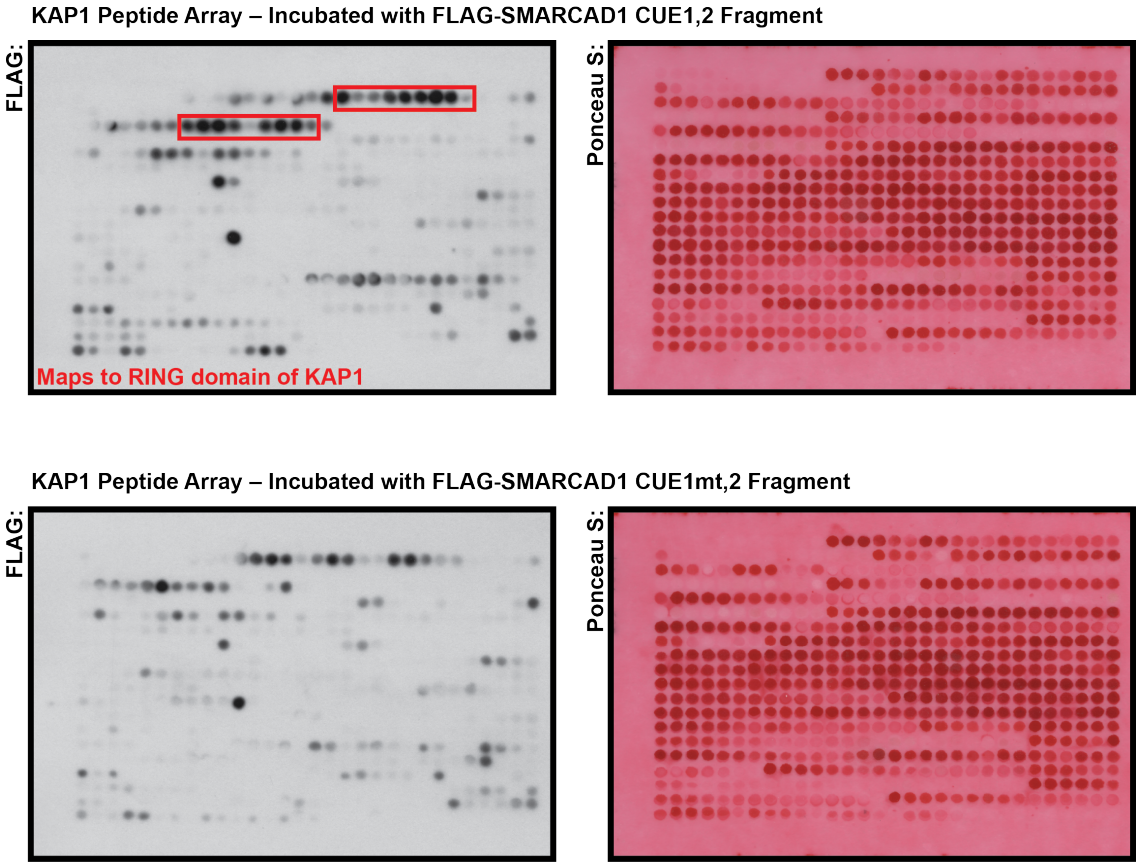


Figure 7.6 – KAP1 Peptide Arrays Cannot Conclusively Reveal the Peptides Recognized by the SMARCAD1 CUE Domain
These peptide arrays array are composed of peptides twenty residues in length that progress along the KAP1 sequence in increments of a single amino acid. Two stretches of peptides interact specifically with the wild type SMARCAD1 CUE domain fragment, but not the CUE1mt,2 fragment, which has a mutated first CUE domain. Both of these clusters of peptides mapped to the RING domain of KAP1.

Two stretches of peptides are recognized specifically by the wild type SMARCAD1 CUE domain fragment, but not the CUE1mt,2 mutant, which has mutations targeting its first CUE domain (Figure 7.6). Both of these areas map to the RING domain of KAP1. It would be judicious to interpret this result with caution since the D202-K434 fragment, which lacks the RING domain, retains some ability to bind the SMARCAD1 CUE domain fragment (Figure 7.4D). Moreover, most RING domains are E3 ubiquitin ligases and the presence of an intrinsic ubiquitin-homology domain within a RING domain has hitherto never been reported (Deshaies and Joazeiro, 2009). In light of the conflicting data, ultimately, it will be necessary to evaluate RING domain KAP1 mutants for the ability to bind SMARCAD1. Additionally, it is possible that the surface of KAP1 recognized by the SMARCAD1 CUE domain is formed by discontiguous amino acids that brought into close proximity upon folding of the protein. Indeed, since conformational epitopes are unrepresented on a peptide array, it is entirely possible that this technique will never reveal the surface with which the SMARCAD1 CUE domain interacts.

7.2.3 Inconclusive Support for a Ubiquitin-Homology Domain in KAP1

The data unanimously indicate that the SMARCAD1-KAP1 interaction strictly depends on the first CUE domain of SMARCAD1. CUE domains are ubiquitin-binding domains with no known ligands apart from ubiquitin and ubiquitin-homology domains. Yet, the interaction between SMARCAD1 and KAP1 can be reconstituted with purified, recombinant proteins that, due to the absence of a functional ubiquitylation system in *E. coli*, are definitely not ubiquitylated, indicating that the residues recognized by the SMARCAD1 CUE domain can intrinsically be located amidst the primary structure of KAP1 (Figure 7.3B). Thus, it was necessary to consider the hypothesis that KAP1 possesses a ubiquitin-homology domain.

If the first CUE domain of SMARCAD1 recognizes a ubiquitin-homology domain within KAP1, it would be expected that an excess of ubiquitin would be able to interfere with the SMARCAD1-KAP1 interaction. However, a hundred fold molar excess of ubiquitin did not affect the ability of the SMARCAD1 CUE domain fragment to interact with KAP1 S33-K434 (Figure 7.7). Thus far, only monomeric

ubiquitin has been tested in this assay; thus, further experiments are required to exclude the possibility that poly-ubiquitin chains are capable of interfering with the SMARCAD1-KAP1 interaction. It is clear, nonetheless, that physiological concentrations of monomeric ubiquitin will not affect the SMARCAD1-KAP1 interaction. Moreover, the inability of ubiquitin to affect the SMARCAD1-KAP1 interaction suggests two hypotheses: 1) the CUE domain of SMARCAD1 may theoretically be able to bind ubiquitin; however, its affinity for ubiquitin significantly lower than that for KAP1, or 2) the first CUE domain of SMARCAD1 recognizes a ligand that is structurally distinct from ubiquitin. For the first scenario to occur, the interaction surface on KAP1 must fill the same hydrophobic groove of the CUE domain in a way that improves binding affinity by more than a hundred fold. This may be possible since the ubiquitin-CUE domain interaction is typically weak, with dissociation constants commonly in the range of a few hundred micromolar.

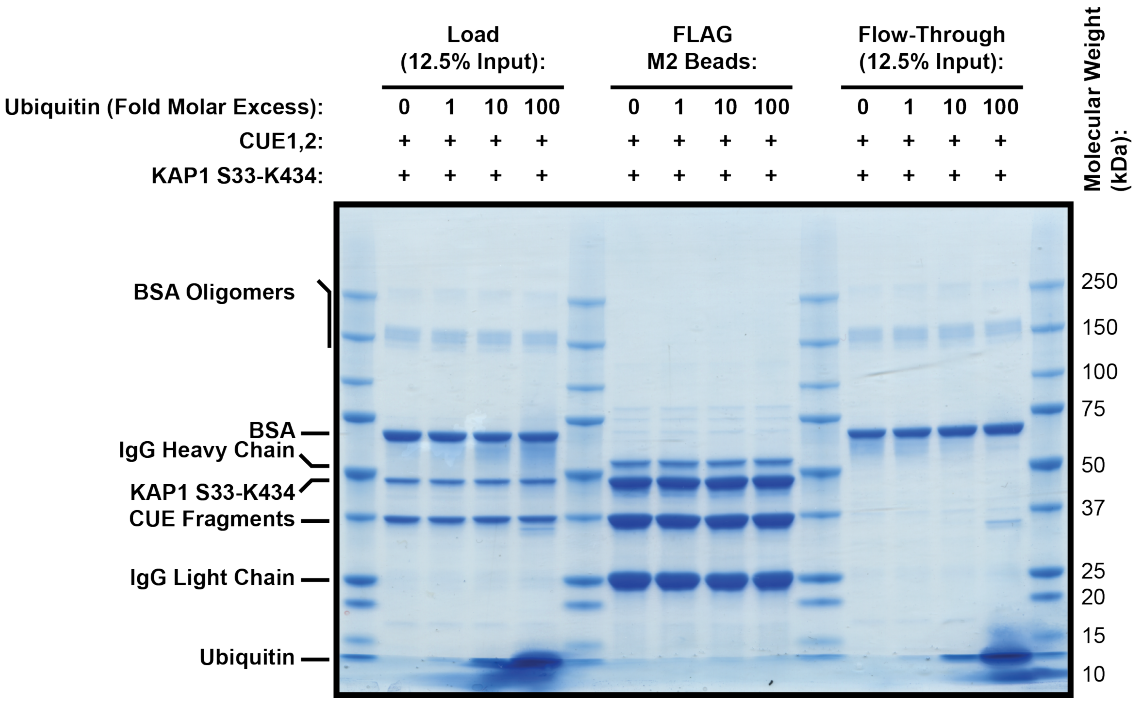


Figure 7.7 – The Interaction between the SMARCAD1 CUE Domain Fragment & KAP1 S33-K434 is Unaffected by an Excess of Ubiquitin
The SMARCAD1 CUE domain fragment was incubated briefly with KAP1 S33-K434 and up to a hundred fold molar excess of ubiquitin. Immunoprecipitation of the CUE domain fragment with FLAG M2 agarose revealed that the excess of ubiquitin did not affect the extent to which the CUE fragment interacted with KAP1 S33-K434.

In silico approaches were then applied to investigate whether any part of KAP1 bore a resemblance to ubiquitin. Comparison of the sequences of ubiquitin and

KAP1 S33-K434 did not reveal any regions of significant homology (Figure 7.8A). Some ubiquitin-homology domains, such as that present in the yeast Ela1 protein, do not possess striking sequence homology to ubiquitin, yet are recognized by a wide range of ubiquitin-binding domains. Despite the limited homology, key ubiquitin residues tend to be conserved or substituted conservatively in *bona fide* ubiquitin-homology domains; for example, in Ela1, the crucial isoleucine 44 residue of ubiquitin was substituted for leucine (Wilson et al., 2013). In contrast, however, key ubiquitin residues were not conserved in the region of KAP1 that bore the greatest sequence similarity to ubiquitin, suggesting that even if a ubiquitin-homology domain were present in KAP1, it is probably not encoded in the primary structure (Figure 7.8A).

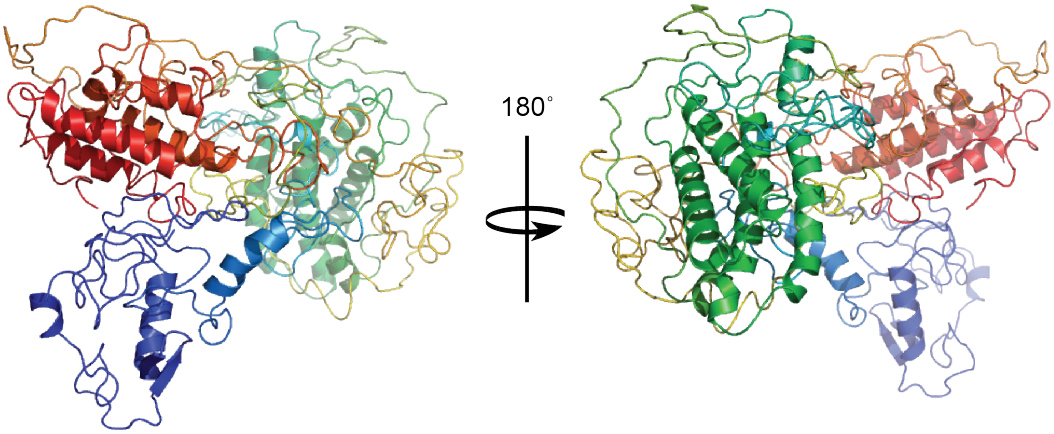
It is possible that a ubiquitin-homology domain is created from the higher order folding of the KAP1 polypeptide. Due to the lack of a reported structure of KAP1, the predicted structure of KAP1 generated by the Phyre2 algorithm was analysed for structural similarity to ubiquitin (Kelley et al., 2015). In this model, 567 residues (68% of the entire sequence) were modelled with accuracy greater than 90%. In fact, specifically considering the RBCC domain, which was of particular interest, 83% of the sequence (333 residues) was modelled with accuracy greater than 90%. Predictably, structures used to construct the model included the coiled-coil domain of TRIM69, and the RING domains of TRIM30 and TRIM5. Additionally, the model was also built upon structures of the human homologue of Ariadne (HHARI), an E3 ubiquitin ligase, and Roquin-1, an RNA-binding zinc-finger protein, as they were also considered to have a high probability of being homologous to KAP1. The predicted KAP1 structure does not bear any similarity to the compact globular ubiquitin fold consisting of a mixture of five β-sheets and an α-helix (Figure 7.8B). Although Phyre2 also performs secondary structure prediction independent of parental structures, this approach similarly failed to identify a region of the RBCC domain with a similar secondary structure configuration to ubiquitin. *A priori* structural prediction is notoriously inaccurate, especially in the absence of structures of multiple homologues. Thus, while suggestive, it certainly cannot be concluded that KAP1 S33-K434 lacks a ubiquitin-homology domain.

A

```
KAP1_S33-K434      STAPSAAASASASAAASSPAGGGAEALELLEHCGVCRERLRPEREPRLLPCLHSACSACL 60
Ubiquitin          ------------------------------------------------------------ 0

KAP1_S33-K434      GPAAPAAANSSGDGGAAGDGTVVDCPVCKQQCFSKDIVENYFMRDSGSKAATDAQDANQC 120
Ubiquitin          ------------------------------------------------------------ 0

KAP1_S33-K434      CTSCEDNAPATSYCVECSEPLCETCVEAHQRVKYTKDHTVRSTGPAKSRDGERTVYCNVH 180
Ubiquitin          ------------------------------------------------------------ 0

KAP1_S33-K434      KHEPLVLFCESCDTLTCRDCQLNAHKDHQYQFLEDAVRNQRKLLASLVKRLGDKHATLQK 240
Ubiquitin          ----------------------------------------MQIFVKTLTGKTITLEV 17
                                                            :  :** *  *  **:

KAP1_S33-K434      STKEVRSSIRQVSDVQKRVQVDVKMAILQIMKELNKRGRVLVNDAQKVTEGQQERLERQH 300
Ubiquitin          EPSDTI------ENVKAKI----------------------QDKEGIPPDQQ----RLI 44
                   . .:.       .:*: ::                     :* : :   **    *

KAP1_S33-K434      WTMTKIQKHQEHILRFASWALESDNNTALLLSKKLIYFQLHRALKMIVDPVEPHGEMKFQ 360
Ubiquitin          FAG-----KQLEDGRTLSDYNIQKESTLHLVL-------RLRG------------------ 75
                   ::     *: *.   ::.: ::.:..  *:*       :*:

KAP1_S33-K434      WDLNAWTKSAEAFGKIVAERPGTNSTGPAPMAPPRAPGPLSK   402
Ubiquitin          ----------------------G--------------------   76
                                       *
```

B

Predicted Structure of KAP1 (Phyre²)



Structure of Ubiquitin (PDB: 1UBQ)

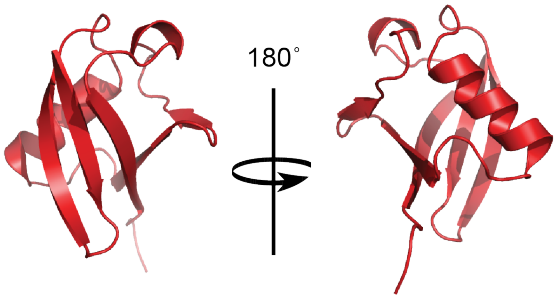


Figure 7.8 – Lack of Sequence or Structural Homology between Ubiquitin and KAP1 S33-K434

A. As assessed by the Clustal Omega algorithm, KAP1 S33-K434 does not contain any regions bearing significant sequence homology to ubiquitin.

B. The predicted structure of KAP1 (Phyre2) is shown above. For comparison, the structure of ubiquitin (PDB: 1UBQ) is shown below. Based on this model, KAP1 does not bear any similarity to the compact globular fold of ubiquitin.

As CUE domains are regarded exclusively as ubiquitin-binding domains, it has been assumed that the CUE domains of SMARCAD1 recognize ligands that structurally resemble ubiquitin. However, beads immobilized with a GST-ubiquitin fusion protein enriched for neither full-length SMARCAD1 nor the SMARCAD1 CUE domain fragment. To favour the interaction between ubiquitin and the SMARCAD1 CUE domains, a high concentration of the purified CUE domain fragment was added to the binding reaction. This was to no avail since the CUE domain fragment did not bind to the GST-ubiquitin fusion protein over background levels, as assessed by the level of non-specific interaction between the CUE domain fragment and the GST-ubiquitin I44A mutant that should suffer from compromised binding.

Empirically, the data suggest that the CUE domains of SMARCAD1 do not recognize ubiquitin. However, it remains formally possible that sub-optimal conditions were used in the *in vitro* binding assay, reducing the interaction between the CUE domains and ubiquitin to below the threshold for detection. Yet, *in silico* approaches indicate a lack of sequence or structural homology between KAP1 and ubiquitin. Therefore, on the balance of probabilities, while unprecedented, the first CUE domain of SMARCAD1 most likely recognizes a region of KAP1 that is structurally distinct from ubiquitin.

7.2.4 Co-Crystallization of the Minimal SMARCAD1-KAP1 Complex

Despite the varied approaches used to characterize the SMARCAD1-KAP1 interaction, its true nature remains inconclusive. This uncertainty can be reduced to one main question – does the region of KAP1 bound by the first CUE domain of SMARCAD1 resemble ubiquitin in terms of its structure? To address this question comprehensively, I am currently attempting to solve the co-crystal structure of the minimal SMARCAD1-KAP1 complex by X-ray crystallography in collaboration with Peter Cherepanov.

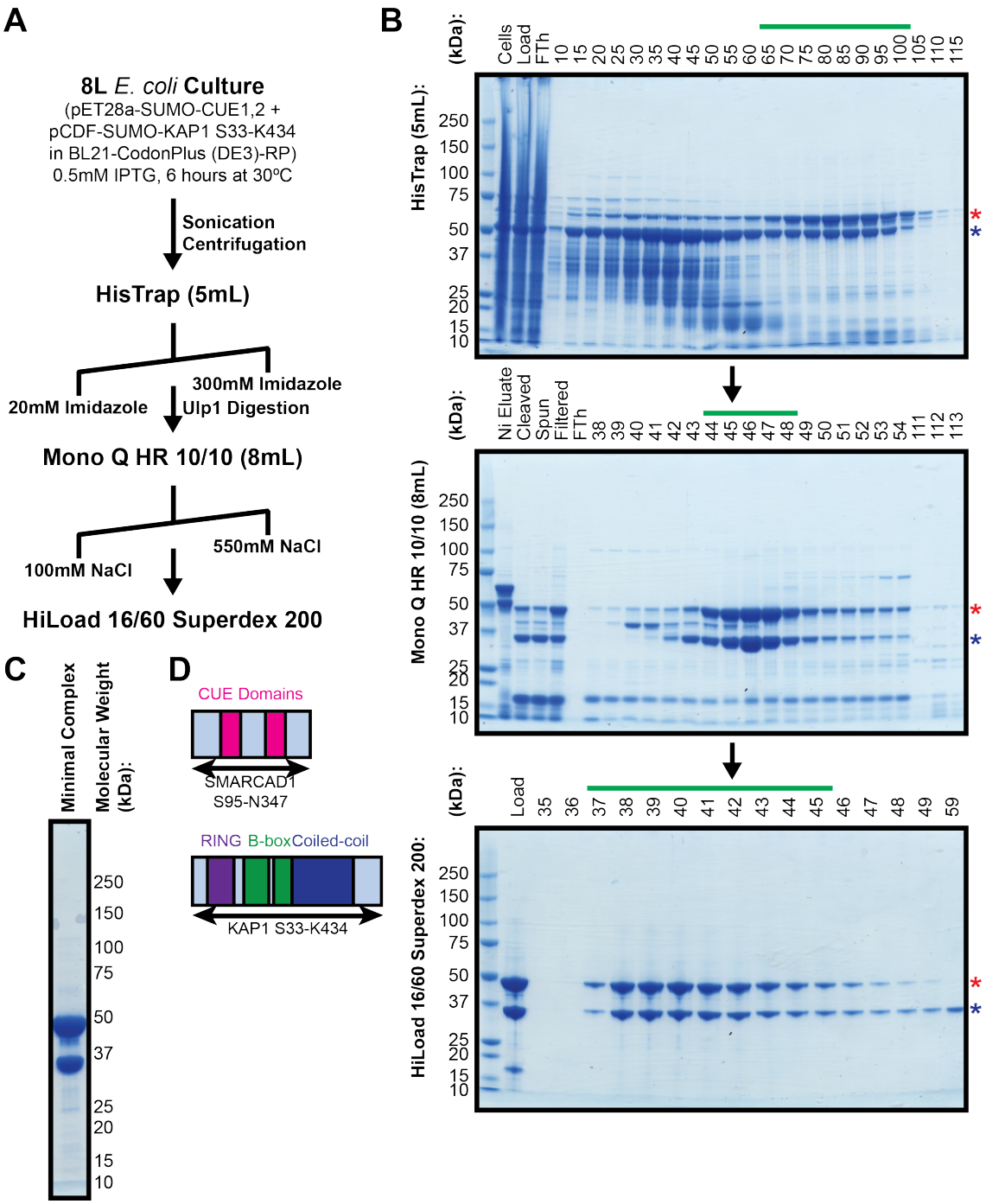


Figure 7.9 – Purification of the Minimal SMARCAD1-KAP1 Complex

A. This schematic outlines the chromatographic steps required to purify the minimal SMARCAD1-KAP1 complex recombinantly co-expressed in *E. coli*.

B. These gels, stained by InstantBlue, from a representative purification of the minimal SMARCAD1-KAP1 complex illustrate the progressive increase in purity over the purification. The green bars above each gel indicate the fractions that were pooled and loaded onto the next column. The red asterisk indicates the KAP1 S33-K434 fragment, while the blue asterisk denotes the SMARCAD1 S95-N346 CUE domain fragment. Interestingly, free CUE domain fragments can be separated from the complex as the latter elutes from the HisTrap column later in the imidazole gradient, suggesting that complex is actually a hetero-multimer with at least two copies of the CUE domain

subunit. The selected fractions from the gel filtration column were pooled and concentrated in spin concentrators to a protein concentration of approximately 15mg/mL.

C. 1µL of the purified minimal SMARCAD1-KAP1 following concentration was resolved by SDS-PAGE and stained with InstantBlue.

D. This schematic illustrates the portions of SMARCAD1 and KAP1 that were co-expressed in *E. coli*.

For this purpose, the minimal SMARCAD1-KAP1 complex was reconstituted by co-expression of the His-tagged, SUMO-CUE domain fragment and untagged SUMO-KAP1 S33-K434 in *E. coli*. It was then purified by sequential affinity, anion exchange and size exclusion chromatography (Figure 7.9A). Interestingly, elution off the initial HisTrap column with an imidazole gradient allows separation of free CUE domain fragments from the minimal SMARCAD1-KAP1 complex, as the latter requires higher imidazole concentrations for elution (Figure 7.9B, top). This suggests that the minimal SMARCAD1-KAP1 complex is actually a hetero-multimer with at least two copies of the CUE domain subunit (since the His tag is only found on the CUE domain fragment). The KAP1 S33-K434 fragment probably dictates oligomerization since its coiled-coil domain is reported to mediate trimerization (Peng et al., 2000). The formal possibility that the enhanced avidity of the complex for nickel is due to a cluster of exposed histidine residues in KAP1 S33-K434 is unlikely since binding of free KAP1 S33-K434 to the nickel column is not observed.

Both the CUE domain fragment and KAP1 S33-K434 co-elute precisely with each other after anion exchange and gel filtration chromatography, indicating that the minimal SMARCAD1-KAP1 complex behaves as a single molecular species, as would be expected for a genuine, stable protein complex (Figure 7.9B, middle & bottom). The minimal SMARCAD1-KAP1 complex is soluble and readily tolerates being concentrated up to 16mg/L with a spin concentrator.

Screening for crystallization conditions for the minimal SMARCAD1-KAP1 complex was performed in 96-well plate format. Fourteen commercial screens were used, allowing 1344 conditions to be tested, though not all the conditions were entirely unique. Small crystals of the SMARCAD1-KAP1 complex were obtained from four conditions (Table 7.1 & Figure 7.10A, C & E). Of the four conditions, three were replicated and could be optimized by further screening to yield larger crystals (Figure 7.10B, D & F). Unfortunately, these crystals diffracted to a resolution of only

~9Å at best, precluding structure determination. Nonetheless, additional constructs are in the process of being tested.

Screen:	Initial (Screen) Condition:	Optimized Condition:
Midas HT-96 (G10)	20% (v/v) Sokalan CP-42 0.1M Lithium acetate 0.1M Bis-Tris NaOH pH 6	19.5% (v/v) Sokalan CP-42 0.1M Lithium acetate 0.1M Bis-Tris NaOH pH5.7 **N.B.** Drop ratio of 2:1 mother liquor to protein
Natrix HT (A3)	10% (v/v) PEG 400 0.2M Potassium chloride 0.01M Magnesium sulphate heptahydrate 0.05M MES NaOH pH 5.6	7.5% (v/v) PEG 400 0.2M Potassium chloride 0.01M Magnesium sulphate heptahydrate 0.05M MES NaOH pH 5.7
JBScreen Kinase HTS (B8)	1M Potassium/Sodium Tartrate 0.1M MES NaOH pH 6.5	1M Potassium/Sodium Tartrate 0.1M MES NaOH pH 6.5

Table 7.1 – Crystallization Conditions for the Minimal SMARCAD1-KAP1 Complex
The conditions used to crystallize the minimal SMARCAD1-KAP1 complex are listed in the table. Unless otherwise stated, drop ratios of 1:1 mother liquor to protein (approximately 15mg/mL) were used. The trays were incubated in a 18°C incubator.

7.3 Conclusions

SMARCAD1 exists in cells as part of a constitutive, stable protein complex with KAP1. Nonetheless, this interaction can be abrogated in cells by point mutations targeting the tandem CUE domains of SMARCAD1. This interaction does not require ubiquitylation of KAP1 since the complex can be reconstituted *in vitro* with purified recombinant proteins that are definitely not ubiquitylated. Mutation and truncation analyses reveal that the interaction occurs between the first CUE domain of SMARCAD1 and KAP1 S33-K434, an N-terminal fragment that spans the entire RBCC domain. The precise nature of the SMARCAD1-KAP1 interaction interface at a structural level remains unclear; on the balance of probabilities, however, the first CUE domain of SMARCAD1 probably recognises a region of KAP1 that lacks structural resemblance of ubiquitin. This is both surprising and interesting. We hope to conclusively address this question by determining the co-crystal structure of a SMARCAD1-KAP1 complex. In addition to clarifying the nature of the SMARCAD1-KAP1 interaction, this structure may offer insights that are generally applicable to our understanding of chromatin and ubiquitin biology.

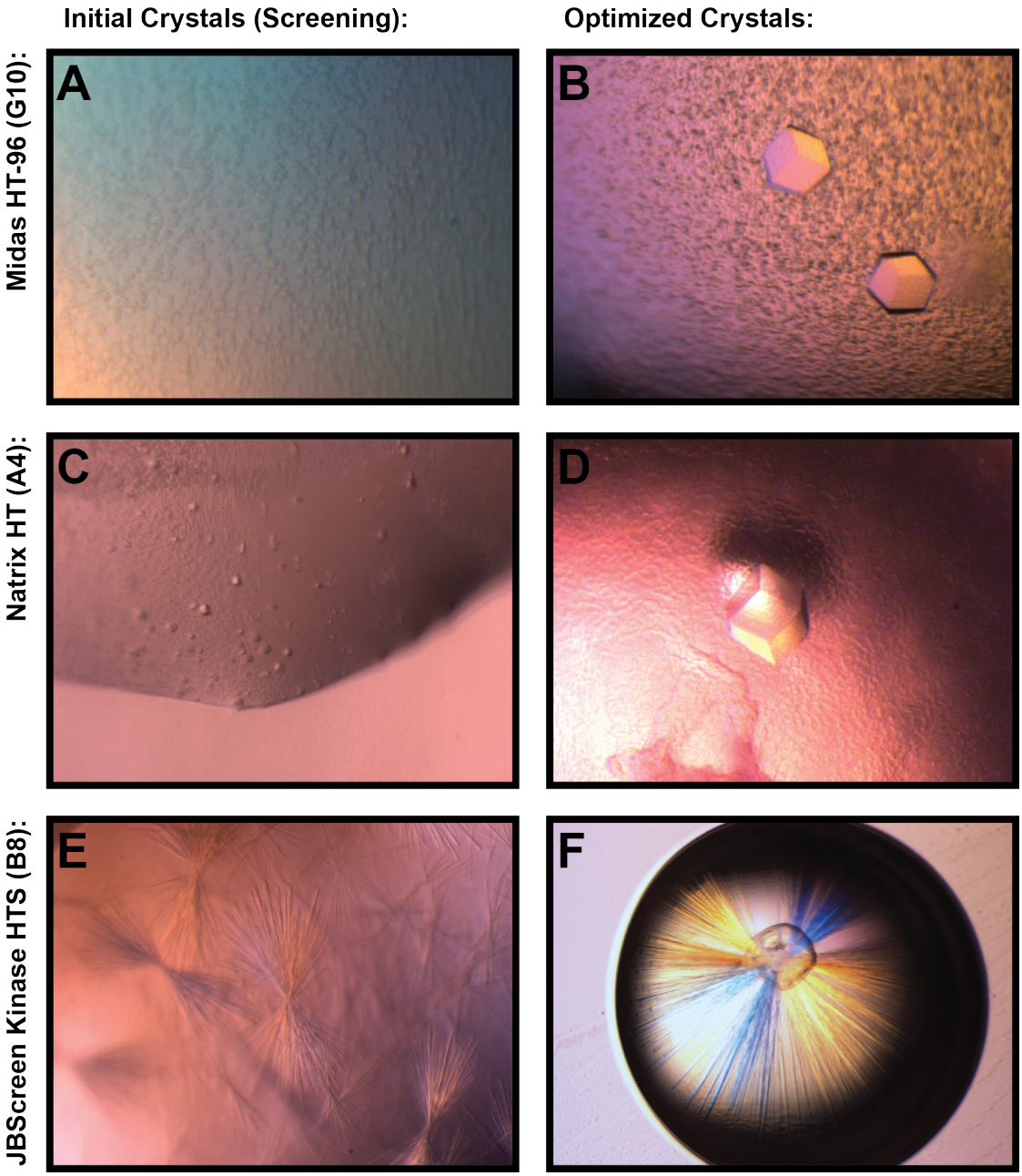


Figure 7.10 – Crystals of the Minimal SMARCAD1-KAP1 Complex
These micrographs depict crystals of the minimal SMARCAD1-KAP1 complex. The crystals on the left (**A**, **C** & **E**) were obtained from conditions represented in commercial crystallization screens. These micrographs were captured under 18.4X magnification, except for **E**, which was magnified 12.8X, though all images were scaled to 15% of their original size. Optimization around the initial conditions in the screens yielded larger crystals, as shown on the right (**B**, **D** & **F**). The magnification of these micrographs is 16X, except for **F**, which was magnified 5.12X; all the images were scaled to 15% of their original sizes.

As discussed in the previous chapter, KAP1 exerts an inhibitory effect on the nucleosome-stimulated ATPase activity of SMARCAD1, implying that the

SMARCAD1-KAP1 interaction is functionally important. Since the SMARCAD1-KAP1 interaction is absolutely dependent on the first CUE domain of SMARCAD1 and the RBCC domain of KAP1, it would be expected that cells expressing only SMARCAD1 bearing point mutations in its first CUE domain would display cellular phenotypes arising from deregulation in SMARCAD1 activity. It is further predicted that the phenotypes observed in these cells would, at least partially, phenocopy those of cells depleted of KAP1. While important, these hypotheses have, however, yet to be tested empirically, and represent an important direction of this project in the future.

Chapter 8. Discussion II: Characterization of the SMARCAD1-KAP1 Chromatin Remodelling Complex

SMARCAD1 is a largely uncharacterized member of the SWI2/SNF2 subfamily of ATPases. Based on sequence homology, it is regarded to be a putative chromatin remodeller. It has been described to exist constitutively as part of a stable protein complex with the heterochromatin-associated protein, KAP1. Functionally, SMARCAD1 promotes the establishment of a repressive chromatin environment, by antagonizing histone hyperacetylation and promoting deposition of the inhibitory H3K9me3 histone mark (Rowbotham et al., 2011). Nevertheless, the SMARCAD1-KAP1 complex has not been characterized at a mechanistic level.

Data presented in this thesis demonstrates for the first time that the SMARCAD1 protein is capable of ATP hydrolysis, especially when stimulated with nucleosomes. Notably, however, the SMARCAD1-KAP1 complex is comparatively less active than the SMARCAD1 ATPase by itself, implicating KAP1 as an inhibitory factor in this context. It is also shown that the interaction between SMARCAD1-KAP1 strictly requires the first CUE domain of SMARCAD1 and KAP1 S33-K434, a region of the KAP1 protein that encompasses its RBCC domain. This interaction does not require ubiquitylation of KAP1 and can be reconstituted *in vitro* with purified, recombinant proteins. Moreover, the data suggests, albeit not definitively, that the first CUE domain recognizes a region of KAP1 that is structurally distinct from ubiquitin. This may, therefore, represent the first report of a novel CUE domain ligand that is unrelated to ubiquitin, though this finding needs to be confirmed by structural studies.

8.1 SMARCAD1 is a Swr1-like Chromatin Remodeller

Members of the Snf2 family of helicase-related proteins can be classified by multiple sequence analysis into twenty-four subfamilies that can be divided into six groups (Flaus et al., 2006). The most well characterized members of the Swr1-like group are Ino80 and Swr1, the catalytic subunits of their eponymous multi-subunit complexes. These proteins are defined by the presence of a long insert between

the two RecA-like lobes of the conserved ATPase domain (Flaus et al., 2006, Clapier and Cairns, 2009). The long inserts are of considerable functional relevance since they serve as an interaction interface through which regulatory subunits associate the catalytic subunit (Wu et al., 2005, Nguyen et al., 2013, Tosi et al., 2013). For example, the inserts of both Swr1 and Ino80 are required for the SWR1 and INO80 chromatin remodelling complexes to associate with their Rvb1/Rvb2 subunits, which themselves, possess helicase activitites. Yet, the structural basis for the interaction differs, not least because SWR1 contains a heterohexameric Rvb1/Rvb2 ring in contrast to the heterododacemeric, stacked Rvb1/Rvb2 component of the INO80 complex (Nguyen et al., 2013, Tosi et al., 2013).

Although SMARCAD1 shares the same overall “split-ATPase” topology as the other members of the Swr1-like family, its insert domain is not as extensive as that of either Swr1 or Ino80. Moreover, SMARCAD1 interacts with neither Rvb1 nor Rvb2 (Rowbotham et al., 2011). It is, nonetheless, important to emphasize that the insert domains mediate stable interactions with other polypeptides other than Rvb1/Rvb2 hetero-oligomers (Wu et al., 2005). Yet, since the only known subunit of the SMARCAD1 chromatin remodelling complex is KAP1, which interacts independently of the insert domain, the significance of the “split-ATPase” topology in SMARCAD1 is unclear. One possibility, albeit entirely speculative, is that the long insert domain of SMARCAD1 mediates transient protein interactions rather than constitutive associations with stable complex subunits.

Recent cryo-EM studies of the SWR1 and INO80 complexes bound to nucleosomes have elucidated the structural basis of nucleosome binding for these two chromatin remodelling complexes (Nguyen et al., 2013, Tosi et al., 2013). In contrast to RSC, another large, multi-subunit chromatin remodelling complex, which accommodates a nucleosome in its central cavity, both SWR1 and INO80 form relatively few contacts with the nucleosome (Chaban et al., 2008, Nguyen et al., 2013, Tosi et al., 2013). Surprisingly, however, the SWR1 and INO80 complexes do not bear much resemblance to each other at a structural level; correspondingly, the basis for their interactions with nucleosomes are entirely idiosyncratic. For example, the nucleosome binds to a depression on the surface of

the SWR1 complex created at an interface between the peripherally located Swr1 subunit and the Rvb1/Rvb2 heterohexamer (Nguyen et al., 2013). In contrast, INO80 forms an elongated, embryo-like structure that cradles a nucleosome by contorting its head-neck-body-foot architecture around it (Tosi et al., 2013). Hence, the dissimilarity in nucleosome binding mechanisms between SWR1 and INO80 prevents extrapolation to predict the structural means by which SMARCAD1 binds to nucleosomes.

A notable difference between SMARCAD1 and other Swr1-like chromatin remodellers is the relative paucity of subunits possessed by SMARCAD1, compared, for instance, to the fourteen and fifteen in the SWR1 and INO80 respectively. Thus, from a structural perspective, it is possible that SMARCAD1 would more closely resemble either the monomeric CHD1 or the heterodimeric ISW1a enzymes. Despite being classified into separate families, both CHD1 and ISWIa possess structurally – though not sequence – conserved SANT and SLIDE DNA binding domains that dramatically enhance the affinity of the enzyme for DNA over the intrinsic binding ability of the ATPase domain (Grune et al., 2003, Ryan et al., 2011). Comparable DNA-binding domains have not been described in either SMARCAD1 or KAP1, suggesting that careful biochemical analysis of the un-annotated regions of SMARCAD1 in search of additional motifs conferring affinity for DNA may represent a profitable line of investigation.

A distinctive feature of SWR1 and the INO80 chromatin remodellers is their ability to catalyse histone exchange (Morrison and Shen, 2009). SWR1 specifically deposits free variant H2AZ/H2B dimers in nucleosomal arrays in place of conventional H2A/H2B dimers (Mizuguchi et al., 2004). INO80 does the exact opposite, substituting incorporated H2AZ/H2B dimers with free H2A/H2B dimers (Papamichos-Chronakis et al., 2011). Interestingly, histone dimer exchange ability has been reported for the yeast homologue of SMARCAD1, Fun30, suggesting that SMARCAD1 may also contribute to histone turnover (Awad et al., 2010). Curiously, depletion of SMARCAD1 results in histone H3 hyperacetylation, a finding that is reminiscent, though the opposite, of the decrease in overall levels of acetylated histone H2AZ in the context of Ino80 deletion (Papamichos-Chronakis et al., 2011, Rowbotham et al., 2011). Taken collectively, it is possible to extend the model

proposed by Rowbotham *et al.*, by speculating that SMARCAD1 is mechanistically linked to the removal of acetylated histone marks from replicated chromatin by promoting histone turnover through dimer exchange.

8.2 Regulation of the SMARCAD1-KAP1 Complex

Although the precise biochemical activities of SMARCAD1 remain to be fully elucidated, its nucleosome-stimulated ATPase activity strongly indicates that it is an active enzyme. More significantly, however, when assembled into a stable complex with SMARCAD1, KAP1 acts as a negative regulator of the ATPase activity of SMARCAD1. KAP1 interacts stoichiometrically with SMARCAD1, suggesting that the majority of SMARCAD1 molecules are kept less active than is within their potential by their constitutive association with KAP1 (Figure 8.1).

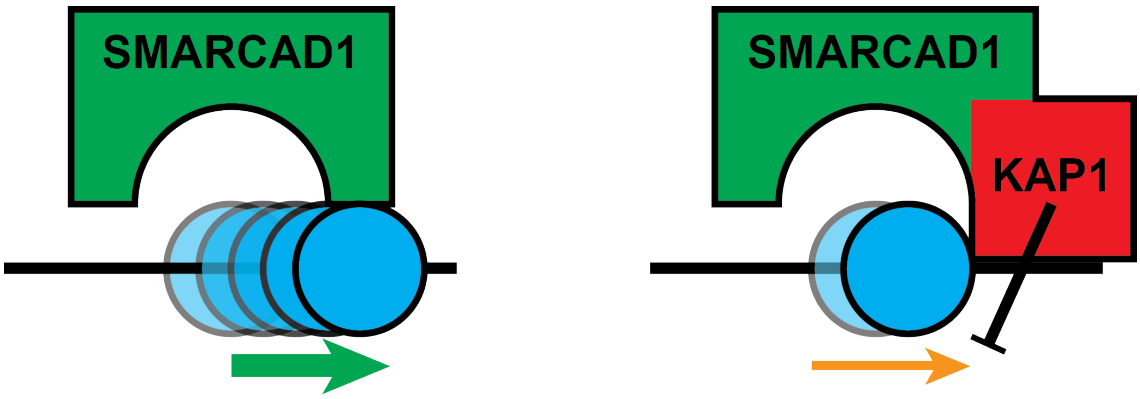


Figure 8.1 – KAP1 is a Negative Regulator of the ATPase Activity of SMARCAD1
By itself, SMARCAD1 displays robust nucleosome-stimulated ATPase activity (left). When in complex with KAP1, however, the nucleosome-stimulated ATPase activity associated with SMARCAD1 is significantly reduced (right). As the main enzymatic activity associated with SMARCAD1 awaits clarification, creative license was adopted to depict SMARCAD1 sliding nucleosomes in this schematic. An important function of KAP may be to maintain the SMARCAD1 ATPase in an inactive state until triggered by a context-specific signal to relieve inhibition, thereby allowing the intrinsically active ATPase domain to act unimpeded.

Maintaining chromatin remodellers in an inactive state by default may be a general regulatory mechanism for cells for several reasons. First, as the human genome encodes for thirty-three proteins that possess a SWI2/SNF2-like helicase domain, it is likely that they are specialized for specific biological processes, with the majority not normally being required. Thus, maintaining a reserve of inactive chromatin remodellers that can be rapidly converted into an active conformation is a simple

yet effective means by which cells can dynamically modify their chromatin structure in response to stimuli. Additionally, unregulated promiscuous ATPase activity from all of the many chromatin remodellers present in each cell may unnecessarily deplete cellular pools of ATP or disrupt normal chromatin structure (Flaus et al., 2006, Narlikar et al., 2013). Indeed, overexpression of the Rhp26 chromatin remodeller in *Schizosacchromyces pombe* resulted in severe cellular toxicity in the absence of exogenous cellular stress. Strikingly, this phenotype was more severe when a hyperactive mutant was overexpressed, suggesting a casual correlation between ATPase hyperactivity and cellular toxicity (Wang et al., 2014a).

There is considerable precedent for chromatin remodellers being maintained in an inactive state. Most notably, the CHD1 ATPase is usually maintained in an open conformation not amenable to ATP hydrolysis. Together with two intervening helices, the two chromodomains form a wedge-like structure – dubbed the 'chromo-wedge' – that occupies the central cleft separating the two recA lobes of the ATPase domain. Moreover, the 'chromo-wedge' physically impedes the second ATPase lobe from rotating the fifty-two degrees required to adopt the closed conformation that is permissible for ATP hydrolysis. In addition, due to the presence of multiple acidic residues, the 'chromo-wedge' is highly negatively charged, allowing it to occupy the DNA-binding surface of the CHD1 ATPase domain (Hauk et al., 2010). It is possible that upon recognition of H3K4me3 by the chromodomains or upon nucleosome binding, the 'chromo-wedge' shifts in position, enabling the CHD1 ATPase lobes to adopt the closed, active conformation.

Similarly, though the ISWI ATPase domain is intrinsically active, it is kept inactive via auto-inhibition by the AutoN regulatory region, which bears sequence similarity with the basic patch of histone H4. It is believed that the AutoN region binds to the ISWI ATPase domains to coerce it to adopt an inactive conformation, in a manner analogous to the CHD1 chromodomains. The histone H4 tail probably competitively displaces AutoN from this interaction surface, thereby allowing ATPase activity and also explaining the requirement for histone H4 for ISWI activity (Clapier and Cairns, 2012). Likewise, the ATPase activity of Rhp26, the *S. pombe* homologue of ERCC6/CSB, is auto-inhibited by a leucine latch motif that binds directly to the ATPase motor, presumably to lock it in an inactive state (Wang et al., 2014a).

Given the ubiquity of auto-inhibition as a regulatory mechanism for chromatin remodellers, it is unsurprising that the SMARCAD1-KAP1 complex is negatively regulated by KAP1. An additional, as yet unknown, stimulatory factor may be required for the complex to become fully active. Though possibilities abound about the identity of this speculative stimulatory factor, one conceptually pleasing hypothesis is that post-translational modification of KAP1 causes a conformational change in the complex that renders the SMARCAD1 ATPase motor fully active. Indeed, KAP1 is not only modified by phosphorylation in an ATM-dependent manner, but also by sumoylation (Ziv et al., 2006, Ivanov et al., 2007, Goodarzi et al., 2011). In this manner, KAP1 may, for instance, act as a signal transformer to specifically activate the SMARCAD1-KAP1 complex following activation of the DNA damage response.

While it is possible to construe a model whereby activation of the SMARCAD1 ATPase occurs when an ubiquitylated substrate competitively displaces KAP1 from the SMARCAD1 CUE domains, such a hypothesis seems quite unlikely. Our data demonstrates that the minimal SMARCAD1-KAP1 complex forms unabated in the presence of a large molar excess of free ubiquitin. Since CUE domains use a shared mechanism for recognizing mono-ubiquitin moieties and poly-ubiquitin chains, those that interact with the former with poor affinity, likewise bind poorly to the latter (Shih et al., 2003, Liu et al., 2012). Hence, while the effect of poly-ubiquitin chains on SMARCAD1-KAP1 complex formation has not been formally tested, it is likely that it too will be similarly inconsequential.

It is also possible that the relatively low activity associated with the SMARCAD1-KAP1 complex is the result of it being stimulated by the wrong substrate. In our assays, we have worked upon the assumption that SMARCAD1 behaves like a typical chromatin remodeller, whose ATPase activity is stimulated by DNA or nucleosomes. It is, therefore, possible that the SMARCAD1-KAP1 is highly active if stimulated with the correct – as yet unknown and untested – substrate. In summary, although it is probably fair to assume that the activity of the SMARCAD1-KAP1 chromatin remodelling complex is tightly regulated by the cell, these regulatory principles await clarification.

8.3 CUE Domains Mediate Protein Interactions

CUE domains are ubiquitin-binding domains that are conserved from yeast to humans (Hicke et al., 2005). Structural studies indicate that each CUE domain is a globular structure comprised of a bundle of three α-helices. Two helices of the CUE domain (α1 and α3), which include the conserved MFP and LL motifs, come into contact with a hydrophobic patch present within the ubiquitin moiety; the interaction is further stabilized by electrostatic contacts along the rim of the interaction interface. Reflecting a relatively limited area of approximately 400Å^2 of intermolecular contacts on each protein, CUE domains typically recognize ubiquitin with only modest affinity (Kang et al., 2003, Prag et al., 2003, Liu et al., 2012). Nevertheless, the dissociation constants for specific CUE domains span a range of approximately 20-160μM (Shih et al., 2003). It has been shown that some CUE domains, such as that of the yeast Vps9p protein, can function either in a monomeric globular state mediating weak binding or as a dimer allowing strong binding – as a dimer, additional contacts are formed between ubiquitin and the typically disengaged, third helix (α2) of the CUE domain (Prag et al., 2003).

Based on sequence conservation, the CUE domains of SMARCAD1 should be capable of binding ubiquitin, though our data, however, suggests otherwise. Yet, the SMARCAD1-KAP1 interaction is abrogated by mutation of the MFP and LL motifs within the first CUE domain of SMARCAD1. The three published structures of CUE domains in complex with ubiquitin demonstrate that while some of these residues form direct hydrophobic interactions with ubiquitin, alanine substitution of the conserved phenylalanine or of the first leucine of the di-leucine repeat motif disrupts proper folding of the CUE domain (Kang et al., 2003, Prag et al., 2003, Liu et al., 2012). On this basis, it is difficult to predict the physical characteristics of the ligand recognized by the first CUE domain, though it is nonetheless likely that the interaction interface is a hydrophobic patch.

The concept of CUE domains mediating protein-protein interactions with non-ubiquitylated ligands is not unprecedented. For example, functioning as a bridging factor, the yeast Def1 protein recruits the Elongin-Cullin (Ela1-Elc1) E3 ubiquitin ligase complex to RNAPII via its CUE domain and a ubiquitin-homology domain in

Ela1. While the ligand in this case was a *bona fide* ubiquitin-homology domain – various ubiquitin-binding domains are capable of interacting with recombinant Ela1-Elc1 – only the Def1 CUE domain was able to enrich for the Elongin-Cullin complex from yeast extracts (Wilson et al., 2013). This observation, therefore, suggests that though the conserved MFP and LL motifs are essential for mediating the interaction interface, CUE domains are also governed by an additional, still unknown, motif that confers ligand specificity. Indeed, the specific enrichment of KAP1 from extract by the SMARCAD1 CUE domains is in accord with this model. Due to this ability for their basal ubiquitin-recognition capability to be tailored for bespoke ligand specificity, CUE domains should be recognized for their capabilities as an interface for protein-protein interactions, rather than as simple ubiquitin-binding domains.

8.4 The KAP1 RBCC Domain as an Interaction Interface

The KAP1 RBCC domain has been previously implicated as a protein interaction surface (Iyengar and Farnham, 2011). Acting as a homotrimer, it interacts with the KRAB repression domain of KRAB domain-containing zinc finger proteins to repress transcription. All three components – the RING domain, tandem B-boxes and coiled-coil domain – are indispensable for the interaction (Friedman et al., 1996, Moosmann et al., 1996, Peng et al., 2000). There is, however, no structural data elucidating the mechanism for this interaction. Undoubtedly a contributing factor is the general resistance to purification presented by fragments of the RBCC domain; for example, extremely low recoveries of the S33-K434 fragment from centrifugal concentrators preclude concentration in that manner. Furthermore, smaller fragments that further truncate the RBCC domain are extremely insoluble.

The recently elucidated structure of the TRIM25 coiled-coil homodimer has offered significant insight into the overall architecture of members of the TRIM family of proteins. Each subunit forms a hairpin structure composed of a short arm – comprised of residues from the linker region that succeeds the coiled-coil domain – that folds back onto a long arm, formed by the coiled-coil domain (Figure 1.16). Dimerization is mediated by the coiled-coil domain and occurs in an antiparallel direction. Uniquely, the coiled-coil domain consists of a combination of heptad (7)

and hendecad (11) amino acid repeats arranged in a symmetrical 7-7-7-7-11-11-11-11-7-7-7-7 pattern, where '7' and '11' refer to a heptad and hendecad repeats respectively. Therefore, a canonical left-handed supercoil is formed by the heptad repeats at either end, whereas the central hendecad repeats generate an under-wound, right-handed coil in the middle. As a consequence of the under-winding, helices from the short arm of the hairpin can interdigitate to pack into a four-helix bundle. Consequently, the domains at the C-terminus of TRIM proteins are centrally positioned on one side of this elongated structure, while the N-terminal RING and B-box domains are located on the apices (Sanchez et al., 2014). It is evident that dimerization and higher order assembly is intrinsic to TRIM proteins. Hence, TRIM proteins probably achieve protein-protein interactions in a bimodal fashion. One category of interactions exploits the modular assembly of TRIM proteins, which enables individual domains to function as discrete structural units that can mediate protein-protein interactions autonomously. The other group of interactions probably involves exposed surfaces created by oligomerization, such as along the coiled-coil domain or the central four-helix bundle.

It was noted that the CUE domain fragment of SMARCAD1 is also able to interact with the D202-K434 KAP1 fragment – which includes only the second B-box and coiled-coil domains – though to a lesser extent than with the entire RBCC domain (S33-K434). This difference in binding is believed to reflect the extremely poor solubility of the D202-K434 fragment. This data suggests, though certainly not conclusively, that the RING and the first B-box domains are redundant for the SMARCAD1-KAP1 interaction. Also, it was noted that a RBCC fragment that excludes the additional linker region (M1-E388) failed to recognize the SMARCAD1 CUE domain fragment, indicating that the short arm of the hairpin is essential for the interaction. KAP1 is probably organized along a similar blueprint to that of TRIM25, since it also conforms to the aforementioned pattern of heptad-hendecad repeats (Sanchez et al., 2014). Taken together, the most likely surface recognized by the SMARCAD1 CUE domain is the four-helix bundle at the centre of the coil.

Our attempts to determine the co-crystal structure of the SMARCAD1 CUE domains in complex with KAP1 S33-K434 by X-ray crystallography are in progress. If obtained, such a structure would undoubtedly be able to definitively clarify our

speculative inferences about the nature of the SMARCAD1-KAP1 interaction. Importantly, it should also offer insights into the biology of KAP1, a pleiotropic chromatin regulator, and also of TRIM proteins in general.

Reference List

ALEN, C., KENT, N. A., JONES, H. S., O'SULLIVAN, J., ARANDA, A. & PROUDFOOT, N. J. 2002. A role for chromatin remodeling in transcriptional termination by RNA polymerase II. *Mol Cell,* 10**,** 1441-52.

AWAD, S., RYAN, D., PROCHASSON, P., OWEN-HUGHES, T. & HASSAN, A. H. 2010. The Snf2 homolog Fun30 acts as a homodimeric ATP-dependent chromatin-remodeling enzyme. *J Biol Chem,* 285**,** 9477-84.

AYGUN, O., SVEJSTRUP, J. & LIU, Y. 2008. A RECQ5-RNA polymerase II association identified by targeted proteomic analysis of human chromatin. *Proc Natl Acad Sci U S A,* 105**,** 8580-4.

AYYANATHAN, K., LECHNER, M. S., BELL, P., MAUL, G. G., SCHULTZ, D. C., YAMADA, Y., TANAKA, K., TORIGOE, K. & RAUSCHER, F. J., 3RD 2003. Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. *Genes Dev,* 17**,** 1855-69.

AZUARA, V., PERRY, P., SAUER, S., SPIVAKOV, M., JORGENSEN, H. F., JOHN, R. M., GOUTI, M., CASANOVA, M., WARNES, G., MERKENSCHLAGER, M. & FISHER, A. G. 2006. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol,* 8**,** 532-8.

BAGCHI, A., PAPAZOGLU, C., WU, Y., CAPURSO, D., BRODT, M., FRANCIS, D., BREDEL, M., VOGEL, H. & MILLS, A. A. 2007. CHD5 is a tumor suppressor at human 1p36. *Cell,* 128**,** 459-75.

BERGS, J. W., NEUENDORFF, N., VAN DER HEIJDEN, G., WASSENAAR, E., REXIN, P., ELSASSER, H. P., MOLL, R., BAARENDS, W. M. & BREHM, A. 2014. Differential expression and sex chromosome association of CHD3/4 and CHD5 during spermatogenesis. *PLoS One,* 9**,** e98203.

BERNSTEIN, B. E., MIKKELSEN, T. S., XIE, X., KAMAL, M., HUEBERT, D. J., CUFF, J., FRY, B., MEISSNER, A., WERNIG, M., PLATH, K., JAENISCH, R., WAGSCHAL, A., FEIL, R., SCHREIBER, S. L. & LANDER, E. S. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell,* 125**,** 315-26.

BIEDERER, T., VOLKWEIN, C. & SOMMER, T. 1997. Role of Cue1p in ubiquitination and degradation at the ER surface. *Science,* 278**,** 1806-9.

BORNENS, M. 1973. Letter: Action of heparin on nuclei: solubilization of chromatin enabling the isolation of nuclear membranes. *Nature,* 244**,** 28-30.

BORNENS, M. 1977. Solubilization of chromatin with heparin and the isolation of nuclear membranes. *Methods Cell Biol,* 15**,** 163-75.

BOYER, L. A., LATEK, R. R. & PETERSON, C. L. 2004. The SANT domain: a unique histone-tail-binding module? *Nat Rev Mol Cell Biol,* 5**,** 158-63.

BOYER, L. A., LEE, T. I., COLE, M. F., JOHNSTONE, S. E., LEVINE, S. S., ZUCKER, J. P., GUENTHER, M. G., KUMAR, R. M., MURRAY, H. L., JENNER, R. G., GIFFORD, D. K., MELTON, D. A., JAENISCH, R. & YOUNG, R. A. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell,* 122**,** 947-56.

CAIRNS, B. R., LORCH, Y., LI, Y., ZHANG, M., LACOMIS, L., ERDJUMENT-BROMAGE, H., TEMPST, P., DU, J., LAURENT, B. & KORNBERG, R. D. 1996. RSC, an essential, abundant chromatin-remodeling complex. *Cell,* 87**,** 1249-60.

CAMMAS, F., HERZOG, M., LEROUGE, T., CHAMBON, P. & LOSSON, R. 2004. Association of the transcriptional corepressor TIF1beta with heterochromatin protein 1 (HP1): an essential role for progression through differentiation. *Genes Dev,* 18**,** 2147-60.

CAMMAS, F., MARK, M., DOLLE, P., DIERICH, A., CHAMBON, P. & LOSSON, R. 2000. Mice lacking the transcriptional corepressor TIF1beta are defective in early postimplantation development. *Development,* 127**,** 2955-63.

CHABAN, Y., EZEOKONKWO, C., CHUNG, W. H., ZHANG, F., KORNBERG, R. D., MAIER-DAVIS, B., LORCH, Y. & ASTURIAS, F. J. 2008. Structure of a RSC-nucleosome complex and insights into chromatin remodeling. *Nat Struct Mol Biol,* 15**,** 1272-7.

CHAKRAVARTY, S., ZENG, L. & ZHOU, M. M. 2009. Structure and site-specific recognition of histone H3 by the PHD finger of human autoimmune regulator. *Structure,* 17**,** 670-9.

CHEN, L., CONAWAY, R. C. & CONAWAY, J. W. 2013. Multiple modes of regulation of the human Ino80 SNF2 ATPase by subunits of the INO80 chromatin-remodeling complex. *Proc Natl Acad Sci U S A,* 110**,** 20497-502.

CHEN, X., CUI, D., PAPUSHA, A., ZHANG, X., CHU, C. D., TANG, J., CHEN, K., PAN, X. & IRA, G. 2012. The Fun30 nucleosome remodeller promotes resection of DNA double-strand break ends. *Nature,* 489**,** 576-80.

CHEN, X., XU, H., YUAN, P., FANG, F., HUSS, M., VEGA, V. B., WONG, E., ORLOV, Y. L., ZHANG, W., JIANG, J., LOH, Y. H., YEO, H. C., YEO, Z. X., NARANG, V., GOVINDARAJAN, K. R., LEONG, B., SHAHAB, A., RUAN, Y., BOURQUE, G., SUNG, W. K., CLARKE, N. D., WEI, C. L. & NG, H. H. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell,* 133**,** 1106-17.

CHRISTOPHOROU, M. A., CASTELO-BRANCO, G., HALLEY-STOTT, R. P., OLIVEIRA, C. S., LOOS, R., RADZISHEUSKAYA, A., MOWEN, K. A., BERTONE, P., SILVA, J. C., ZERNICKA-GOETZ, M., NIELSEN, M. L., GURDON, J. B. & KOUZARIDES, T. 2014. Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature,* 507**,** 104-8.

CLAPIER, C. R. & CAIRNS, B. R. 2009. The biology of chromatin remodeling complexes. *Annu Rev Biochem,* 78**,** 273-304.

CLAPIER, C. R. & CAIRNS, B. R. 2012. Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature,* 492**,** 280-4.

CORONA, D. F., LANGST, G., CLAPIER, C. R., BONTE, E. J., FERRARI, S., TAMKUN, J. W. & BECKER, P. B. 1999. ISWI is an ATP-dependent nucleosome remodeling factor. *Mol Cell,* 3**,** 239-45.

COSTELLOE, T., LOUGE, R., TOMIMATSU, N., MUKHERJEE, B., MARTINI, E., KHADAROO, B., DUBOIS, K., WIEGANT, W. W., THIERRY, A., BURMA, S., VAN ATTIKUM, H. & LLORENTE, B. 2012. The yeast Fun30 and human SMARCAD1 chromatin remodellers promote DNA end resection. *Nature,* 489**,** 581-4.

COURVALIN, J. C., DUMONTIER, M. & BORNENS, M. 1982. Solubilization of nuclear structures by the polyanion heparin. *J Biol Chem,* 257**,** 456-63.

DANG, W. & BARTHOLOMEW, B. 2007. Domain architecture of the catalytic subunit in the ISW2-nucleosome complex. *Mol Cell Biol,* 27**,** 8306-17.

DEINDL, S., HWANG, W. L., HOTA, S. K., BLOSSER, T. R., PRASAD, P., BARTHOLOMEW, B. & ZHUANG, X. 2013. ISWI remodelers slide nucleosomes with coordinated multi-base-pair entry steps and single-base-pair exit steps. *Cell,* 152**,** 442-52.

DESHAIES, R. J. & JOAZEIRO, C. A. 2009. RING domain E3 ubiquitin ligases. *Annu Rev Biochem,* 78**,** 399-434.

DURR, H., KORNER, C., MULLER, M., HICKMANN, V. & HOPFNER, K. P. 2005. X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. *Cell,* 121**,** 363-73.

EFRONI, S., DUTTAGUPTA, R., CHENG, J., DEHGHANI, H., HOEPPNER, D. J., DASH, C., BAZETT-JONES, D. P., LE GRICE, S., MCKAY, R. D., BUETOW, K. H., GINGERAS, T. R., MISTELI, T. & MESHORER, E. 2008. Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell,* 2**,** 437-47.

EGAN, C. M., NYMAN, U., SKOTTE, J., STREUBEL, G., TURNER, S., O'CONNELL, D. J., RRAKLLI, V., DOLAN, M. J., CHADDERTON, N., HANSEN, K., FARRAR, G. J., HELIN, K., HOLMBERG, J. & BRACKEN, A. P. 2013. CHD5 is required for neurogenesis and has a dual role in facilitating gene expression and polycomb gene repression. *Dev Cell,* 26**,** 223-36.

EHRENSBERGER, A. H. & KORNBERG, R. D. 2011. Isolation of an activator-dependent, promoter-specific chromatin remodeling factor. *Proc Natl Acad Sci U S A,* 108**,** 10115-20.

ESKELAND, R., LEEB, M., GRIMES, G. R., KRESS, C., BOYLE, S., SPROUL, D., GILBERT, N., FAN, Y., SKOULTCHI, A. I., WUTZ, A. & BICKMORE, W. A. 2010. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell,* 38**,** 452-64.

FILIPPAKOPOULOS, P., PICAUD, S., MANGOS, M., KEATES, T., LAMBERT, J. P., BARSYTE-LOVEJOY, D., FELLETAR, I., VOLKMER, R., MULLER, S.,

PAWSON, T., GINGRAS, A. C., ARROWSMITH, C. H. & KNAPP, S. 2012. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell,* 149**,** 214-31.

FLANAGAN, J. F., MI, L. Z., CHRUSZCZ, M., CYMBOROWSKI, M., CLINES, K. L., KIM, Y., MINOR, W., RASTINEJAD, F. & KHORASANIZADEH, S. 2005. Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature,* 438**,** 1181-5.

FLAUS, A., MARTIN, D. M., BARTON, G. J. & OWEN-HUGHES, T. 2006. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res,* 34**,** 2887-905.

FLOTHO, A. & MELCHIOR, F. 2013. Sumoylation: a regulatory protein modification in health and disease. *Annu Rev Biochem,* 82**,** 357-85.

FRASER, R. A., HEARD, D. J., ADAM, S., LAVIGNE, A. C., LE DOUARIN, B., TORA, L., LOSSON, R., ROCHETTE-EGLY, C. & CHAMBON, P. 1998. The putative cofactor TIF1alpha is a protein kinase that is hyperphosphorylated upon interaction with liganded nuclear receptors. *J Biol Chem,* 273**,** 16199-204.

FRIEDMAN, J. R., FREDERICKS, W. J., JENSEN, D. E., SPEICHER, D. W., HUANG, X. P., NEILSON, E. G. & RAUSCHER, F. J., 3RD 1996. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev,* 10**,** 2067-78.

FUJITA, N., JAYE, D. L., KAJITA, M., GEIGERMAN, C., MORENO, C. S. & WADE, P. A. 2003. MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer. *Cell,* 113**,** 207-19.

FUJITA, T., IGARASHI, J., OKAWA, E. R., GOTOH, T., MANNE, J., KOLLA, V., KIM, J., ZHAO, H., PAWEL, B. R., LONDON, W. B., MARIS, J. M., WHITE, P. S. & BRODEUR, G. M. 2008. CHD5, a tumor suppressor gene deleted from 1p36.31 in neuroblastomas. *J Natl Cancer Inst,* 100**,** 940-9.

GANTKE, T., BOUSSOUF, S., JANZEN, J., MORRICE, N. A., HOWELL, S., MUHLBERGER, E. & LEY, S. C. 2013. Ebola virus VP35 induces high-level production of recombinant TPL-2-ABIN-2-NF-kappaB1 p105 complex in co-transfected HEK-293 cells. *Biochem J,* 452**,** 359-65.

GARCIA-CAO, M., O'SULLIVAN, R., PETERS, A. H., JENUWEIN, T. & BLASCO, M. A. 2004. Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nat Genet,* 36**,** 94-9.

GASPAR-MAIA, A., ALAJEM, A., POLESSO, F., SRIDHARAN, R., MASON, M. J., HEIDERSBACH, A., RAMALHO-SANTOS, J., MCMANUS, M. T., PLATH, K., MESHORER, E. & RAMALHO-SANTOS, M. 2009. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature,* 460**,** 863-8.

GERHOLD, C. B. & GASSER, S. M. 2014. INO80 and SWR complexes: relating structure to function in chromatin remodeling. *Trends Cell Biol,* 24**,** 619-31.

GIFFORD, C. A., ZILLER, M. J., GU, H., TRAPNELL, C., DONAGHEY, J., TSANKOV, A., SHALEK, A. K., KELLEY, D. R., SHISHKIN, A. A., ISSNER, R., ZHANG, X., COYNE, M., FOSTEL, J. L., HOLMES, L., MELDRIM, J., GUTTMAN, M., EPSTEIN, C., PARK, H., KOHLBACHER, O., RINN, J., GNIRKE, A., LANDER, E. S., BERNSTEIN, B. E. & MEISSNER, A. 2013. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell,* 153**,** 1149-63.

GKIKOPOULOS, T., SCHOFIELD, P., SINGH, V., PINSKAYA, M., MELLOR, J., SMOLLE, M., WORKMAN, J. L., BARTON, G. J. & OWEN-HUGHES, T. 2011. A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. *Science,* 333**,** 1758-60.

GNATT, A., FU, J. & KORNBERG, R. D. 1997. Formation and crystallization of yeast RNA polymerase II elongation complexes. *J Biol Chem,* 272**,** 30799-805.

GOODARZI, A. A., KURKA, T. & JEGGO, P. A. 2011. KAP-1 phosphorylation regulates CHD3 nucleosome remodeling during the DNA double-strand break response. *Nat Struct Mol Biol,* 18**,** 831-9.

GOODARZI, A. A., NOON, A. T., DECKBAR, D., ZIV, Y., SHILOH, Y., LOBRICH, M. & JEGGO, P. A. 2008. ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Mol Cell,* 31**,** 167-77.

GRABBE, C. & DIKIC, I. 2009. Functional roles of ubiquitin-like domain (ULD) and ubiquitin-binding domain (UBD) containing proteins. *Chem Rev,* 109**,** 1481-94.

GRAU, D. J., CHAPMAN, B. A., GARLICK, J. D., BOROWSKY, M., FRANCIS, N. J. & KINGSTON, R. E. 2011. Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev,* 25**,** 2210-21.

GREWAL, S. I. & MOAZED, D. 2003. Heterochromatin and epigenetic control of gene expression. *Science,* 301**,** 798-802.

GRUNE, T., BRZESKI, J., EBERHARTER, A., CLAPIER, C. R., CORONA, D. F., BECKER, P. B. & MULLER, C. W. 2003. Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol Cell,* 12**,** 449-60.

GUNTHER, K., RUST, M., LEERS, J., BOETTGER, T., SCHARFE, M., JAREK, M., BARTKUHN, M. & RENKAWITZ, R. 2013. Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. *Nucleic Acids Res,* 41**,** 3010-21.

HALL, I. M., SHANKARANARAYANA, G. D., NOMA, K., AYOUB, N., COHEN, A. & GREWAL, S. I. 2002. Establishment and maintenance of a heterochromatin domain. *Science,* 297**,** 2232-7.

HATAKEYAMA, S. 2011. TRIM proteins and cancer. *Nat Rev Cancer,* 11**,** 792-804.

HAUK, G., MCKNIGHT, J. N., NODELMAN, I. M. & BOWMAN, G. D. 2010. The chromodomains of the Chd1 chromatin remodeler regulate DNA access to the ATPase motor. *Mol Cell,* 39**,** 711-23.

HAWKINS, R. D., HON, G. C., LEE, L. K., NGO, Q., LISTER, R., PELIZZOLA, M., EDSALL, L. E., KUAN, S., LUU, Y., KLUGMAN, S., ANTOSIEWICZ-BOURGET, J., YE, Z., ESPINOZA, C., AGARWAHL, S., SHEN, L., RUOTTI, V., WANG, W., STEWART, R., THOMSON, J. A., ECKER, J. R. & REN, B. 2010. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell,* 6**,** 479-91.

HEITZ, E. 1928. *Das Heterochromatin der Moose*, Bornträger.

HENDRICH, B., GUY, J., RAMSAHOYE, B., WILSON, V. A. & BIRD, A. 2001. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev,* 15**,** 710-23.

HICKE, L., SCHUBERT, H. L. & HILL, C. P. 2005. Ubiquitin-binding domains. *Nat Rev Mol Cell Biol,* 6**,** 610-21.

HO, L., JOTHI, R., RONAN, J. L., CUI, K., ZHAO, K. & CRABTREE, G. R. 2009a. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Natl Acad Sci U S A,* 106**,** 5187-91.

HO, L., MILLER, E. L., RONAN, J. L., HO, W. Q., JOTHI, R. & CRABTREE, G. R. 2011. esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat Cell Biol,* 13**,** 903-13.

HO, L., RONAN, J. L., WU, J., STAAHL, B. T., CHEN, L., KUO, A., LESSARD, J., NESVIZHSKII, A. I., RANISH, J. & CRABTREE, G. R. 2009b. An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A,* 106**,** 5181-6.

HOTA, S. K., BHARDWAJ, S. K., DEINDL, S., LIN, Y. C., ZHUANG, X. & BARTHOLOMEW, B. 2013. Nucleosome mobilization by ISW2 requires the concerted action of the ATPase and SLIDE domains. *Nat Struct Mol Biol,* 20**,** 222-9.

HU, G., KIM, J., XU, Q., LENG, Y., ORKIN, S. H. & ELLEDGE, S. J. 2009. A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev,* 23**,** 837-48.

HU, X., MALIK, S., NEGROIU, C. C., HUBBARD, K., VELALAR, C. N., HAMPTON, B., GROSU, D., CATALANO, J., ROEDER, R. G. & GNATT, A. 2006. A Mediator-responsive form of metazoan RNA polymerase II. *Proc Natl Acad Sci U S A,* 103**,** 9506-11.

HUSNJAK, K. & DIKIC, I. 2012. Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. *Annu Rev Biochem,* 81**,** 291-322.

HWANG, W. L., DEINDL, S., HARADA, B. T. & ZHUANG, X. 2014. Histone H4 tail mediates allosteric regulation of nucleosome remodelling by linker DNA. *Nature,* 512**,** 213-7.

ITO, T., LEVENSTEIN, M. E., FYODOROV, D. V., KUTACH, A. K., KOBAYASHI, R. & KADONAGA, J. T. 1999. ACF consists of two subunits, Acf1 and ISWI, that

function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev,* 13**,** 1529-39.

IVANOV, A. V., PENG, H., YURCHENKO, V., YAP, K. L., NEGOREV, D. G., SCHULTZ, D. C., PSULKOWSKI, E., FREDERICKS, W. J., WHITE, D. E., MAUL, G. G., SADOFSKY, M. J., ZHOU, M. M. & RAUSCHER, F. J., 3RD 2007. PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol Cell,* 28**,** 823-37.

IYENGAR, S. & FARNHAM, P. J. 2011. KAP1 protein: an enigmatic master regulator of the genome. *J Biol Chem,* 286**,** 26267-76.

IYENGAR, S., IVANOV, A. V., JIN, V. X., RAUSCHER, F. J., 3RD & FARNHAM, P. J. 2011. Functional analysis of KAP1 genomic recruitment. *Mol Cell Biol,* 31**,** 1833-47.

IYER, V. & STRUHL, K. 1995. Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J,* 14**,** 2570-9.

JACOBS, S. A. & KHORASANIZADEH, S. 2002. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science,* 295**,** 2080-3.

JASIN, M. & ROTHSTEIN, R. 2013. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol,* 5**,** a012740.

JASON, L. J., FINN, R. M., LINDSEY, G. & AUSIO, J. 2005. Histone H2A ubiquitination does not preclude histone H1 binding, but it facilitates its association with the nucleosome. *J Biol Chem,* 280**,** 4975-82.

KADOCH, C. & CRABTREE, G. R. 2013. Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma. *Cell,* 153**,** 71-85.

KAJI, K., CABALLERO, I. M., MACLEOD, R., NICHOLS, J., WILSON, V. A. & HENDRICH, B. 2006. The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat Cell Biol,* 8**,** 285-92.

KANG, R. S., DANIELS, C. M., FRANCIS, S. A., SHIH, S. C., SALERNO, W. J., HICKE, L. & RADHAKRISHNAN, I. 2003. Solution structure of a CUE-ubiquitin complex reveals a conserved mode of ubiquitin binding. *Cell,* 113**,** 621-30.

KAPLAN, N., MOORE, I. K., FONDUFE-MITTENDORF, Y., GOSSETT, A. J., TILLO, D., FIELD, Y., LEPROUST, E. M., HUGHES, T. R., LIEB, J. D., WIDOM, J. & SEGAL, E. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature,* 458**,** 362-6.

KELLEY, D. E., STOKES, D. G. & PERRY, R. P. 1999. CHD1 interacts with SSRP1 and depends on both its chromodomain and its ATPase/helicase-like domain for proper association with chromatin. *Chromosoma,* 108**,** 10-25.

KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc,* 10**,** 845-58.

KOLLA, V., NARAPARAJU, K., ZHUANG, T., HIGASHI, M., KOLLA, S., BLOBEL, G. A. & BRODEUR, G. M. 2015. The tumour suppressor CHD5 forms a NuRD-type chromatin remodelling complex. *Biochem J,* 468**,** 345-52.

KOLLA, V., ZHUANG, T., HIGASHI, M., NARAPARAJU, K. & BRODEUR, G. M. 2014. Role of CHD5 in human cancers: 10 years later. *Cancer Res,* 74**,** 652-8.

KOMANDER, D. & RAPE, M. 2012. The ubiquitin code. *Annu Rev Biochem,* 81**,** 203-29.

KORECKA, J. A., VAN KESTEREN, R. E., BLAAS, E., SPITZER, S. O., KAMSTRA, J. H., SMIT, A. B., SWAAB, D. F., VERHAAGEN, J. & BOSSERS, K. 2013. Phenotypic characterization of retinoic acid differentiated SH-SY5Y cells by transcriptional profiling. *PLoS One,* 8**,** e63862.

KORNBERG, R. D. 1974. Chromatin structure: a repeating unit of histones and DNA. *Science,* 184**,** 868-71.

KOUZARIDES, T. 2007. Chromatin modifications and their function. *Cell,* 128**,** 693-705.

KROGAN, N. J., KIM, M., AHN, S. H., ZHONG, G., KOBOR, M. S., CAGNEY, G., EMILI, A., SHILATIFARD, A., BURATOWSKI, S. & GREENBLATT, J. F. 2002. RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. *Mol Cell Biol,* 22**,** 6979-92.

KULAEVA, O. I., NIZOVTSEVA, E. V., POLIKANOV, Y. S., ULIANOV, S. V. & STUDITSKY, V. M. 2012. Distant activation of transcription: mechanisms of enhancer action. *Mol Cell Biol,* 32**,** 4892-7.

LAI, A. Y. & WADE, P. A. 2011. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat Rev Cancer,* 11**,** 588-96.

LANTERMANN, A. B., STRAUB, T., STRALFORS, A., YUAN, G. C., EKWALL, K. & KORBER, P. 2010. Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. *Nat Struct Mol Biol,* 17**,** 251-7.

LAUBERTH, S. M., NAKAYAMA, T., WU, X., FERRIS, A. L., TANG, Z., HUGHES, S. H. & ROEDER, R. G. 2013. H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell,* 152**,** 1021-36.

LE DOUARIN, B., NIELSEN, A. L., GARNIER, J. M., ICHINOSE, H., JEANMOUGIN, F., LOSSON, R. & CHAMBON, P. 1996. A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors. *EMBO J,* 15**,** 6701-15.

LE GUEZENNEC, X., VERMEULEN, M., BRINKMAN, A. B., HOEIJMAKERS, W. A., COHEN, A., LASONDER, E. & STUNNENBERG, H. G. 2006. MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol Cell Biol,* 26**,** 843-51.

LECHNER, M. S., BEGG, G. E., SPEICHER, D. W. & RAUSCHER, F. J., 3RD 2000. Molecular determinants for targeting heterochromatin protein 1-mediated gene

silencing: direct chromoshadow domain-KAP-1 corepressor interaction is essential. *Mol Cell Biol,* 20**,** 6449-65.

LESSARD, J., WU, J. I., RANISH, J. A., WAN, M., WINSLOW, M. M., STAAHL, B. T., WU, H., AEBERSOLD, R., GRAEF, I. A. & CRABTREE, G. R. 2007. An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron,* 55**,** 201-15.

LI, W., WU, J., KIM, S. Y., ZHAO, M., HEARN, S. A., ZHANG, M. Q., MEISTRICH, M. L. & MILLS, A. A. 2014. Chd5 orchestrates chromatin remodelling during sperm development. *Nat Commun,* 5**,** 3812.

LI, Z., SCHUG, J., TUTEJA, G., WHITE, P. & KAESTNER, K. H. 2011. The nucleosome map of the mammalian liver. *Nat Struct Mol Biol,* 18**,** 742-6.

LIANG, Q., DENG, H., LI, X., WU, X., TANG, Q., CHANG, T. H., PENG, H., RAUSCHER, F. J., 3RD, OZATO, K. & ZHU, F. 2011. Tripartite motif-containing protein 28 is a small ubiquitin-related modifier E3 ligase and negative regulator of IFN regulatory factor 7. *J Immunol,* 187**,** 4754-63.

LICKERT, H., TAKEUCHI, J. K., VON BOTH, I., WALLS, J. R., MCAULIFFE, F., ADAMSON, S. L., HENKELMAN, R. M., WRANA, J. L., ROSSANT, J. & BRUNEAU, B. G. 2004. Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature,* 432**,** 107-12.

LIN, J. J., LEHMANN, L. W., BONORA, G., SRIDHARAN, R., VASHISHT, A. A., TRAN, N., PLATH, K., WOHLSCHLEGEL, J. A. & CAREY, M. 2011. Mediator coordinates PIC assembly with recruitment of CHD1. *Genes Dev,* 25**,** 2198-209.

LIU, S., CHEN, Y., LI, J., HUANG, T., TARASOV, S., KING, A., WEISSMAN, A. M., BYRD, R. A. & DAS, R. 2012. Promiscuous interactions of gp78 E3 ligase CUE domain with polyubiquitin chains. *Structure,* 20**,** 2138-50.

LORCH, Y., LAPOINTE, J. W. & KORNBERG, R. D. 1987. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell,* 49**,** 203-10.

LORCH, Y., MAIER-DAVIS, B. & KORNBERG, R. D. 2006. Chromatin remodeling by nucleosome disassembly in vitro. *Proc Natl Acad Sci U S A,* 103**,** 3090-3.

LORCH, Y., ZHANG, M. & KORNBERG, R. D. 1999. Histone octamer transfer by a chromatin-remodeling complex. *Cell,* 96**,** 389-92.

LUGER, K., DECHASSA, M. L. & TREMETHICK, D. J. 2012. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol,* 13**,** 436-47.

LUGER, K., MADER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature,* 389**,** 251-60.

LUSSER, A., URWIN, D. L. & KADONAGA, J. T. 2005. Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nat Struct Mol Biol,* 12**,** 160-6.

MANSFIELD, R. E., MUSSELMAN, C. A., KWAN, A. H., OLIVER, S. S., GARSKE, A. L., DAVRAZOU, F., DENU, J. M., KUTATELADZE, T. G. & MACKAY, J. P. 2011. Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. *J Biol Chem,* 286**,** 11779-91.

MARBLESTONE, J. G., EDAVETTAL, S. C., LIM, Y., LIM, P., ZUO, X. & BUTT, T. R. 2006. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci,* 15**,** 182-9.

MARFELLA, C. G. & IMBALZANO, A. N. 2007. The Chd family of chromatin remodelers. *Mutat Res,* 618**,** 30-40.

MASSIAH, M. A., SIMMONS, B. N., SHORT, K. M. & COX, T. C. 2006. Solution structure of the RBCC/TRIM B-box1 domain of human MID1: B-box with a RING. *J Mol Biol,* 358**,** 532-45.

MAVRICH, T. N., JIANG, C., IOSHIKHES, I. P., LI, X., VENTERS, B. J., ZANTON, S. J., TOMSHO, L. P., QI, J., GLASER, R. L., SCHUSTER, S. C., GILMOUR, D. S., ALBERT, I. & PUGH, B. F. 2008. Nucleosome organization in the Drosophila genome. *Nature,* 453**,** 358-62.

MCKNIGHT, J. N., JENKINS, K. R., NODELMAN, I. M., ESCOBAR, T. & BOWMAN, G. D. 2011. Extranucleosomal DNA binding directs nucleosome sliding by Chd1. *Mol Cell Biol,* 31**,** 4746-59.

MESHORER, E., YELLAJOSHULA, D., GEORGE, E., SCAMBLER, P. J., BROWN, D. T. & MISTELI, T. 2006. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell,* 10**,** 105-16.

MIKKELSEN, T. S., KU, M., JAFFE, D. B., ISSAC, B., LIEBERMAN, E., GIANNOUKOS, G., ALVAREZ, P., BROCKMAN, W., KIM, T. K., KOCHE, R. P., LEE, W., MENDENHALL, E., O'DONOVAN, A., PRESSER, A., RUSS, C., XIE, X., MEISSNER, A., WERNIG, M., JAENISCH, R., NUSBAUM, C., LANDER, E. S. & BERNSTEIN, B. E. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature,* 448**,** 553-60.

MIN, J., ZHANG, Y. & XU, R. M. 2003. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev,* 17**,** 1823-8.

MIZUGUCHI, G., SHEN, X., LANDRY, J., WU, W. H., SEN, S. & WU, C. 2004. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science,* 303**,** 343-8.

MOHRMANN, L. & VERRIJZER, C. P. 2005. Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochim Biophys Acta,* 1681**,** 59-73.

MOOSMANN, P., GEORGIEV, O., LE DOUARIN, B., BOURQUIN, J. P. & SCHAFFNER, W. 1996. Transcriptional repression by RING finger protein TIF1 beta that interacts with the KRAB repressor domain of KOX1. *Nucleic Acids Res,* 24**,** 4859-67.

MORRA, R., LEE, B. M., SHAW, H., TUMA, R. & MANCINI, E. J. 2012. Concerted action of the PHD, chromo and motor domains regulates the human chromatin remodelling ATPase CHD4. *FEBS Lett,* 586**,** 2513-21.

MORRIS, S. A., BAEK, S., SUNG, M. H., JOHN, S., WIENCH, M., JOHNSON, T. A., SCHILTZ, R. L. & HAGER, G. L. 2014. Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. *Nat Struct Mol Biol,* 21**,** 73-81.

MORRISON, A. J. & SHEN, X. 2009. Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. *Nat Rev Mol Cell Biol,* 10**,** 373-84.

MOSHKIN, Y. M., CHALKLEY, G. E., KAN, T. W., REDDY, B. A., OZGUR, Z., VAN IJCKEN, W. F., DEKKERS, D. H., DEMMERS, J. A., TRAVERS, A. A. & VERRIJZER, C. P. 2012. Remodelers organize cellular chromatin by counteracting intrinsic histone-DNA sequence preferences in a class-specific manner. *Mol Cell Biol,* 32**,** 675-88.

MUELLER-PLANITZ, F., KLINKER, H., LUDWIGSEN, J. & BECKER, P. B. 2013. The ATPase domain of ISWI is an autonomous nucleosome remodeling machine. *Nat Struct Mol Biol,* 20**,** 82-9.

MURAMATSU, D., SINGH, P. B., KIMURA, H., TACHIBANA, M. & SHINKAI, Y. 2013. Pericentric heterochromatin generated by HP1 protein interaction-defective histone methyltransferase Suv39h1. *J Biol Chem,* 288**,** 25285-96.

MUSSELMAN, C. A., AVVAKUMOV, N., WATANABE, R., ABRAHAM, C. G., LALONDE, M. E., HONG, Z., ALLEN, C., ROY, S., NUNEZ, J. K., NICKOLOFF, J., KULESZA, C. A., YASUI, A., COTE, J. & KUTATELADZE, T. G. 2012a. Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. *Nat Struct Mol Biol,* 19**,** 1266-72.

MUSSELMAN, C. A., RAMIREZ, J., SIMS, J. K., MANSFIELD, R. E., OLIVER, S. S., DENU, J. M., MACKAY, J. P., WADE, P. A., HAGMAN, J. & KUTATELADZE, T. G. 2012b. Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression. *Proc Natl Acad Sci U S A,* 109**,** 787-92.

NAKAYAMA, J., RICE, J. C., STRAHL, B. D., ALLIS, C. D. & GREWAL, S. I. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science,* 292**,** 110-3.

NARLIKAR, G. J., SUNDARAMOORTHY, R. & OWEN-HUGHES, T. 2013. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell,* 154**,** 490-503.

NGUYEN, V. Q., RANJAN, A., STENGEL, F., WEI, D., AEBERSOLD, R., WU, C. & LESCHZINER, A. E. 2013. Molecular architecture of the ATP-dependent chromatin-remodeling complex SWR1. *Cell,* 154**,** 1220-31.

NIELSEN, A. L., ORTIZ, J. A., YOU, J., OULAD-ABDELGHANI, M., KHECHUMIAN, R., GANSMULLER, A., CHAMBON, P. & LOSSON, R. 1999. Interaction with members of the heterochromatin protein 1 (HP1) family and histone

deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. *EMBO J,* 18**,** 6385-95.

NODELMAN, I. M. & BOWMAN, G. D. 2013. Nucleosome sliding by Chd1 does not require rigid coupling between DNA-binding and ATPase domains. *EMBO Rep,* 14**,** 1098-103.

O'GEEN, H., SQUAZZO, S. L., IYENGAR, S., BLAHNIK, K., RINN, J. L., CHANG, H. Y., GREEN, R. & FARNHAM, P. J. 2007. Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. *PLoS Genet,* 3**,** e89.

OKAWA, E. R., GOTOH, T., MANNE, J., IGARASHI, J., FUJITA, T., SILVERMAN, K. A., XHAO, H., MOSSE, Y. P., WHITE, P. S. & BRODEUR, G. M. 2008. Expression and sequence analysis of candidates for the 1p36.31 tumor suppressor gene deleted in neuroblastomas. *Oncogene,* 27**,** 803-10.

OLIVER, S. S., MUSSELMAN, C. A., SRINIVASAN, R., SVAREN, J. P., KUTATELADZE, T. G. & DENU, J. M. 2012. Multivalent recognition of histone tails by the PHD fingers of CHD5. *Biochemistry,* 51**,** 6534-44.

ORUETXEBARRIA, I., VENTURINI, F., KEKARAINEN, T., HOUWELING, A., ZUIJDERDUIJN, L. M., MOHD-SARIP, A., VRIES, R. G., HOEBEN, R. C. & VERRIJZER, C. P. 2004. P16INK4a is required for hSNF5 chromatin remodeler-induced cellular senescence in malignant rhabdoid tumor cells. *J Biol Chem,* 279**,** 3807-16.

OWEN, D. J., ORNAGHI, P., YANG, J. C., LOWE, N., EVANS, P. R., BALLARIO, P., NEUHAUS, D., FILETICI, P. & TRAVERS, A. A. 2000. The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *EMBO J,* 19**,** 6141-9.

PANIER, S. & DUROCHER, D. 2013. Push back to respond better: regulatory inhibition of the DNA double-strand break response. *Nat Rev Mol Cell Biol,* 14**,** 661-72.

PAPAMICHOS-CHRONAKIS, M., WATANABE, S., RANDO, O. J. & PETERSON, C. L. 2011. Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell,* 144**,** 200-13.

PATEL, A., MCKNIGHT, J. N., GENZOR, P. & BOWMAN, G. D. 2011. Identification of residues in chromodomain helicase DNA-binding protein 1 (Chd1) required for coupling ATP hydrolysis to nucleosome sliding. *J Biol Chem,* 286**,** 43984-93.

PATEL, D. J. & WANG, Z. 2013. Readout of epigenetic modifications. *Annu Rev Biochem,* 82**,** 81-118.

PAUL, S., KUO, A., SCHALCH, T., VOGEL, H., JOSHUA-TOR, L., MCCOMBIE, W. R., GOZANI, O., HAMMELL, M. & MILLS, A. A. 2013. Chd5 requires PHD-mediated histone 3 binding for tumor suppression. *Cell Rep,* 3**,** 92-102.

PENG, H., BEGG, G. E., SCHULTZ, D. C., FRIEDMAN, J. R., JENSEN, D. E., SPEICHER, D. W. & RAUSCHER, F. J., 3RD 2000. Reconstitution of the KRAB-KAP-1 repressor complex: a model system for defining the molecular

anatomy of RING-B box-coiled-coil domain-mediated protein-protein interactions. *J Mol Biol,* 295**,** 1139-62.

PETERS, A. H., KUBICEK, S., MECHTLER, K., O'SULLIVAN, R. J., DERIJCK, A. A., PEREZ-BURGOS, L., KOHLMAIER, A., OPRAVIL, S., TACHIBANA, M., SHINKAI, Y., MARTENS, J. H. & JENUWEIN, T. 2003. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell,* 12**,** 1577-89.

PIATTI, P., LIM, C. Y., NAT, R., VILLUNGER, A., GELEY, S., SHUE, Y. T., SORATROI, C., MOSER, M. & LUSSER, A. 2015. Embryonic stem cell differentiation requires full length Chd1. *Sci Rep,* 5**,** 8007.

POINTNER, J., PERSSON, J., PRASAD, P., NORMAN-AXELSSON, U., STRALFORS, A., KHOROSJUTINA, O., KRIETENSTEIN, N., SVENSSON, J. P., EKWALL, K. & KORBER, P. 2012. CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in S. pombe. *EMBO J,* 31**,** 4388-403.

PONTING, C. P. 2000. Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction. *Biochem J,* 351 Pt 2**,** 527-35.

POTTS, R. C., ZHANG, P., WURSTER, A. L., PRECHT, P., MUGHAL, M. R., WOOD, W. H., 3RD, ZHANG, Y., BECKER, K. G., MATTSON, M. P. & PAZIN, M. J. 2011. CHD5, a brain-specific paralog of Mi2 chromatin remodeling enzymes, regulates expression of neuronal genes. *PLoS One,* 6**,** e24515.

PRAG, G., MISRA, S., JONES, E. A., GHIRLANDO, R., DAVIES, B. A., HORAZDOVSKY, B. F. & HURLEY, J. H. 2003. Mechanism of ubiquitin recognition by the CUE domain of Vps9p. *Cell,* 113**,** 609-20.

QIU, Y., LIU, L., ZHAO, C., HAN, C., LI, F., ZHANG, J., WANG, Y., LI, G., MEI, Y., WU, M., WU, J. & SHI, Y. 2012. Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ reveals a regulatory mechanism for HOXA9 transcription. *Genes Dev,* 26**,** 1376-91.

QUAN, J., ADELMANT, G., MARTO, J. A., LOOK, A. T. & YUSUFZAI, T. 2014. The chromatin remodeling factor CHD5 is a transcriptional repressor of WEE1. *PLoS One,* 9**,** e108066.

QUAN, J. & YUSUFZAI, T. 2014. The tumor suppressor chromodomain helicase DNA-binding protein 5 (CHD5) remodels nucleosomes by unwrapping. *J Biol Chem,* 289**,** 20717-26.

RAMIREZ, J., DEGE, C., KUTATELADZE, T. G. & HAGMAN, J. 2012. MBD2 and multiple domains of CHD4 are required for transcriptional repression by Mi-2/NuRD complexes. *Mol Cell Biol,* 32**,** 5078-88.

RANJAN, A., MIZUGUCHI, G., FITZGERALD, P. C., WEI, D., WANG, F., HUANG, Y., LUK, E., WOODCOCK, C. L. & WU, C. 2013. Nucleosome-free region dominates histone acetylation in targeting SWR1 to promoters for H2A.Z replacement. *Cell,* 154**,** 1232-45.

ROBINSON, P. J., FAIRALL, L., HUYNH, V. A. & RHODES, D. 2006. EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc Natl Acad Sci U S A,* 103**,** 6506-11.

ROWBOTHAM, S. P., BARKI, L., NEVES-COSTA, A., SANTOS, F., DEAN, W., HAWKES, N., CHOUDHARY, P., WILL, W. R., WEBSTER, J., OXLEY, D., GREEN, C. M., VARGA-WEISZ, P. & MERMOUD, J. E. 2011. Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. *Mol Cell,* 42**,** 285-96.

RYAN, D. P., SUNDARAMOORTHY, R., MARTIN, D., SINGH, V. & OWEN-HUGHES, T. 2011. The DNA-binding domain of the Chd1 chromatin-remodelling enzyme contains SANT and SLIDE domains. *EMBO J,* 30**,** 2596-609.

RYAN, R. F., SCHULTZ, D. C., AYYANATHAN, K., SINGH, P. B., FRIEDMAN, J. R., FREDERICKS, W. J. & RAUSCHER, F. J., 3RD 1999. KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Kruppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. *Mol Cell Biol,* 19**,** 4366-78.

SAHA, A., WITTMEYER, J. & CAIRNS, B. R. 2002. Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes Dev,* 16**,** 2120-34.

SAHA, A., WITTMEYER, J. & CAIRNS, B. R. 2005. Chromatin remodeling through directional DNA translocation from an internal nucleosomal site. *Nat Struct Mol Biol,* 12**,** 747-55.

SANCHEZ, J. G., OKREGLICKA, K., CHANDRASEKARAN, V., WELKER, J. M., SUNDQUIST, W. I. & PORNILLOS, O. 2014. The tripartite motif coiled-coil is an elongated antiparallel hairpin dimer. *Proc Natl Acad Sci U S A,* 111**,** 2494-9.

SANTOS-ROSA, H., SCHNEIDER, R., BANNISTER, A. J., SHERRIFF, J., BERNSTEIN, B. E., EMRE, N. C., SCHREIBER, S. L., MELLOR, J. & KOUZARIDES, T. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature,* 419**,** 407-11.

SCHALCH, T., DUDA, S., SARGENT, D. F. & RICHMOND, T. J. 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature,* 436**,** 138-41.

SCHONES, D. E., CUI, K., CUDDAPAH, S., ROH, T. Y., BARSKI, A., WANG, Z., WEI, G. & ZHAO, K. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell,* 132**,** 887-98.

SCHULTZ, D. C., AYYANATHAN, K., NEGOREV, D., MAUL, G. G. & RAUSCHER, F. J., 3RD 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev,* 16**,** 919-32.

SCHULTZ, D. C., FRIEDMAN, J. R. & RAUSCHER, F. J., 3RD 2001. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes Dev,* 15**,** 428-43.

SCHWARTZ, Y. B. & PIRROTTA, V. 2013. A new world of Polycombs: unexpected partnerships and emerging functions. *Nat Rev Genet,* 14**,** 853-64.

SEGAL, E., FONDUFE-MITTENDORF, Y., CHEN, L., THASTROM, A., FIELD, Y., MOORE, I. K., WANG, J. P. & WIDOM, J. 2006. A genomic code for nucleosome positioning. *Nature,* 442**,** 772-8.

SHAHBAZIAN, M. D. & GRUNSTEIN, M. 2007. Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem,* 76**,** 75-100.

SHARMA, A., JENKINS, K. R., HEROUX, A. & BOWMAN, G. D. 2011. Crystal structure of the chromodomain helicase DNA-binding protein 1 (Chd1) DNA-binding domain in complex with DNA. *J Biol Chem,* 286**,** 42099-104.

SHIH, S. C., PRAG, G., FRANCIS, S. A., SUTANTO, M. A., HURLEY, J. H. & HICKE, L. 2003. A ubiquitin-binding motif required for intramolecular monoubiquitylation, the CUE domain. *EMBO J,* 22**,** 1273-81.

SHIM, Y. S., CHOI, Y., KANG, K., CHO, K., OH, S., LEE, J., GREWAL, S. I. & LEE, D. 2012. Hrp3 controls nucleosome positioning to suppress non-coding transcription in eu- and heterochromatin. *EMBO J,* 31**,** 4375-87.

SHLYUEVA, D., STAMPFEL, G. & STARK, A. 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet,* 15**,** 272-86.

SIMIC, R., LINDSTROM, D. L., TRAN, H. G., ROINICK, K. L., COSTA, P. J., JOHNSON, A. D., HARTZOG, G. A. & ARNDT, K. M. 2003. Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *EMBO J,* 22**,** 1846-56.

SIMS, R. J., 3RD, CHEN, C. F., SANTOS-ROSA, H., KOUZARIDES, T., PATEL, S. S. & REINBERG, D. 2005. Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J Biol Chem,* 280**,** 41789-92.

SIMS, R. J., 3RD, MILLHOUSE, S., CHEN, C. F., LEWIS, B. A., ERDJUMENT-BROMAGE, H., TEMPST, P., MANLEY, J. L. & REINBERG, D. 2007. Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol Cell,* 28**,** 665-76.

SKENE, P. J., HERNANDEZ, A. E., GROUDINE, M. & HENIKOFF, S. 2014. The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. *Elife,* 3**,** e02042.

SMOLLE, M., VENKATESH, S., GOGOL, M. M., LI, H., ZHANG, Y., FLORENS, L., WASHBURN, M. P. & WORKMAN, J. L. 2012. Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat Struct Mol Biol,* 19**,** 884-92.

SOBEL, R. E., COOK, R. G., PERRY, C. A., ANNUNZIATO, A. T. & ALLIS, C. D. 1995. Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4. *Proc Natl Acad Sci U S A,* 92**,** 1237-41.

SONG, F., CHEN, P., SUN, D., WANG, M., DONG, L., LIANG, D., XU, R. M., ZHU, P. & LI, G. 2014. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science,* 344**,** 376-80.

SRIPATHY, S. P., STEVENS, J. & SCHULTZ, D. C. 2006. The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Mol Cell Biol,* 26**,** 8623-38.

STOCKDALE, C., FLAUS, A., FERREIRA, H. & OWEN-HUGHES, T. 2006. Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. *J Biol Chem,* 281**,** 16279-88.

STOKES, D. G., TARTOF, K. D. & PERRY, R. P. 1996. CHD1 is concentrated in interbands and puffed regions of Drosophila polytene chromosomes. *Proc Natl Acad Sci U S A,* 93**,** 7137-42.

STRUHL, K. & SEGAL, E. 2013. Determinants of nucleosome positioning. *Nat Struct Mol Biol,* 20**,** 267-73.

SUGANUMA, T. & WORKMAN, J. L. 2011. Signals and combinatorial functions of histone modifications. *Annu Rev Biochem,* 80**,** 473-99.

SUKA, N., LUO, K. & GRUNSTEIN, M. 2002. Sir2p and Sas2p opposingly regulate acetylation of yeast histone H4 lysine16 and spreading of heterochromatin. *Nat Genet,* 32**,** 378-83.

SVEJSTRUP, J. Q., WANG, Z., FEAVER, W. J., WU, X., BUSHNELL, D. A., DONAHUE, T. F., FRIEDBERG, E. C. & KORNBERG, R. D. 1995. Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome. *Cell,* 80**,** 21-8.

SZERLONG, H., HINATA, K., VISWANATHAN, R., ERDJUMENT-BROMAGE, H., TEMPST, P. & CAIRNS, B. R. 2008. The HSA domain binds nuclear actin-related proteins to regulate chromatin-remodeling ATPases. *Nat Struct Mol Biol,* 15**,** 469-76.

TAKAHASHI, K., TANABE, K., OHNUKI, M., NARITA, M., ICHISAKA, T., TOMODA, K. & YAMANAKA, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell,* 131**,** 861-72.

TAO, H., SIMMONS, B. N., SINGIREDDY, S., JAKKIDI, M., SHORT, K. M., COX, T. C. & MASSIAH, M. A. 2008. Structure of the MID1 tandem B-boxes reveals an interaction reminiscent of intermolecular ring heterodimers. *Biochemistry,* 47**,** 2450-7.

TEIF, V. B., VAINSHTEIN, Y., CAUDRON-HERGER, M., MALLM, J. P., MARTH, C., HOFER, T. & RIPPE, K. 2012. Genome-wide nucleosome positioning during embryonic stem cell development. *Nat Struct Mol Biol,* 19**,** 1185-92.

TESSARZ, P. & KOUZARIDES, T. 2014. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol,* 15**,** 703-8.

THOMPSON, N. E., ARONSON, D. B. & BURGESS, R. R. 1990. Purification of eukaryotic RNA polymerase II by immunoaffinity chromatography. Elution of active enzyme with protein stabilizing agents from a polyol-responsive monoclonal antibody. *J Biol Chem,* 265**,** 7069-77.

TONG, J. K., HASSIG, C. A., SCHNITZLER, G. R., KINGSTON, R. E. & SCHREIBER, S. L. 1998. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature,* 395**,** 917-21.

TORRENTE, M. P., ZEE, B. M., YOUNG, N. L., BALIBAN, R. C., LEROY, G., FLOUDAS, C. A., HAKE, S. B. & GARCIA, B. A. 2011. Proteomic interrogation of human chromatin. *PLoS One,* 6**,** e24747.

TOSI, A., HAAS, C., HERZOG, F., GILMOZZI, A., BERNINGHAUSEN, O., UNGEWICKELL, C., GERHOLD, C. B., LAKOMEK, K., AEBERSOLD, R., BECKMANN, R. & HOPFNER, K. P. 2013. Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell,* 154**,** 1207-19.

TRAN, H. G., STEGER, D. J., IYER, V. R. & JOHNSON, A. D. 2000. The chromo domain protein chd1p from budding yeast is an ATP-dependent chromatin-modifying factor. *EMBO J,* 19**,** 2323-31.

TROJER, P. & REINBERG, D. 2007. Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell,* 28**,** 1-13.

TSAI, W. W., WANG, Z., YIU, T. T., AKDEMIR, K. C., XIA, W., WINTER, S., TSAI, C. Y., SHI, X., SCHWARZER, D., PLUNKETT, W., ARONOW, B., GOZANI, O., FISCHLE, W., HUNG, M. C., PATEL, D. J. & BARTON, M. C. 2010. TRIM24 links a non-canonical histone signature to breast cancer. *Nature,* 468**,** 927-32.

TSUKIYAMA, T., PALMER, J., LANDEL, C. C., SHILOACH, J. & WU, C. 1999. Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodeling factors in Saccharomyces cerevisiae. *Genes Dev,* 13**,** 686-97.

VALOUEV, A., JOHNSON, S. M., BOYD, S. D., SMITH, C. L., FIRE, A. Z. & SIDOW, A. 2011. Determinants of nucleosome organization in primary human cells. *Nature,* 474**,** 516-20.

VAN DER KNAAP, J. A., KOZHEVNIKOVA, E., LANGENBERG, K., MOSHKIN, Y. M. & VERRIJZER, C. P. 2010. Biosynthetic enzyme GMP synthetase cooperates with ubiquitin-specific protease 7 in transcriptional regulation of ecdysteroid target genes. *Mol Cell Biol,* 30**,** 736-44.

VAN DER KNAAP, J. A., KUMAR, B. R., MOSHKIN, Y. M., LANGENBERG, K., KRIJGSVELD, J., HECK, A. J., KARCH, F. & VERRIJZER, C. P. 2005. GMP synthetase stimulates histone H2B deubiquitylation by the epigenetic silencer USP7. *Mol Cell,* 17**,** 695-707.

VERMEULEN, M., MULDER, K. W., DENISSOV, S., PIJNAPPEL, W. W., VAN SCHAIK, F. M., VARIER, R. A., BALTISSEN, M. P., STUNNENBERG, H. G., MANN, M. & TIMMERS, H. T. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell,* 131**,** 58-69.

VERSTEEGE, I., SEVENET, N., LANGE, J., ROUSSEAU-MERCK, M. F., AMBROS, P., HANDGRETINGER, R., AURIAS, A. & DELATTRE, O. 1998. Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature,* 394**,** 203-6.

VIJAY-KUMAR, S., BUGG, C. E. & COOK, W. J. 1987. Structure of ubiquitin refined at 1.8 A resolution. *J Mol Biol,* 194**,** 531-44.

VOIGT, P., TEE, W. W. & REINBERG, D. 2013. A double take on bivalent promoters. *Genes Dev,* 27**,** 1318-38.

VRIES, R. G., BEZROOKOVE, V., ZUIJDERDUIJN, L. M., KIA, S. K., HOUWELING, A., ORUETXEBARRIA, I., RAAP, A. K. & VERRIJZER, C. P. 2005. Cancer-associated mutations in chromatin remodeler hSNF5 promote chromosomal instability by compromising the mitotic checkpoint. *Genes Dev,* 19**,** 665-70.

WADE, P. A., GEGONNE, A., JONES, P. L., BALLESTAR, E., AUBRY, F. & WOLFFE, A. P. 1999. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat Genet,* 23**,** 62-6.

WANG, H. B. & ZHANG, Y. 2001. Mi2, an auto-antigen for dermatomyositis, is an ATP-dependent nucleosome remodeling factor. *Nucleic Acids Res,* 29**,** 2517-21.

WANG, J., RAO, S., CHU, J., SHEN, X., LEVASSEUR, D. N., THEUNISSEN, T. W. & ORKIN, S. H. 2006. A protein interaction network for pluripotency of embryonic stem cells. *Nature,* 444**,** 364-8.

WANG, L., LIMBO, O., FEI, J., CHEN, L., KIM, B., LUO, J., CHONG, J., CONAWAY, R. C., CONAWAY, J. W., RANISH, J. A., KADONAGA, J. T., RUSSELL, P. & WANG, D. 2014a. Regulation of the Rhp26ERCC6/CSB chromatin remodeler by a novel conserved leucine latch motif. *Proc Natl Acad Sci U S A,* 111**,** 18566-71.

WANG, Y., CHAKRAVARTY, P., RANES, M., KELLY, G., BROOKS, P. J., NEILAN, E., STEWART, A., SCHIAVO, G. & SVEJSTRUP, J. Q. 2014b. Dysregulation of gene expression as a cause of Cockayne syndrome neurological disease. *Proc Natl Acad Sci U S A,* 111**,** 14454-9.

WATANABE, S., RADMAN-LIVAJA, M., RANDO, O. J. & PETERSON, C. L. 2013. A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. *Science,* 340**,** 195-9.

WATANABE, S., RESCH, M., LILYESTROM, W., CLARK, N., HANSEN, J. C., PETERSON, C. & LUGER, K. 2010. Structural characterization of H3K56Q nucleosomes and nucleosomal arrays. *Biochim Biophys Acta,* 1799**,** 480-6.

WATSON, A. A., MAHAJAN, P., MERTENS, H. D., DEERY, M. J., ZHANG, W., PHAM, P., DU, X., BARTKE, T., ZHANG, W., EDLICH, C., BERRIDGE, G., CHEN, Y., BURGESS-BROWN, N. A., KOUZARIDES, T., WIECHENS, N., OWEN-HUGHES, T., SVERGUN, D. I., GILEADI, O. & LAUE, E. D. 2012. The PHD and chromo domains regulate the ATPase activity of the human chromatin remodeler CHD4. *J Mol Biol,* 422**,** 3-17.

WEAKE, V. M. & WORKMAN, J. L. 2008. Histone ubiquitination: triggering gene activity. *Mol Cell,* 29**,** 653-63.

WEST, J. A., COOK, A., ALVER, B. H., STADTFELD, M., DEATON, A. M., HOCHEDLINGER, K., PARK, P. J., TOLSTORUKOV, M. Y. & KINGSTON, R. E. 2014. Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nat Commun,* 5**,** 4719.

WILSON, M. D., HARREMAN, M., TASCHNER, M., REID, J., WALKER, J., ERDJUMENT-BROMAGE, H., TEMPST, P. & SVEJSTRUP, J. Q. 2013. Proteasome-mediated processing of Def1, a critical step in the cellular response to transcription stress. *Cell,* 154**,** 983-95.

WILSON, M. D., SAPONARO, M., LEIDL, M. A. & SVEJSTRUP, J. Q. 2012. MultiDsk: a ubiquitin-specific affinity resin. *PLoS One,* 7**,** e46398.

WOOD, A., SCHNEIDER, J., DOVER, J., JOHNSTON, M. & SHILATIFARD, A. 2003. The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *J Biol Chem,* 278**,** 34739-42.

WOODCOCK, C. L. & GHOSH, R. P. 2010. Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol,* 2**,** a000596.

WU, J. I., LESSARD, J., OLAVE, I. A., QIU, Z., GHOSH, A., GRAEF, I. A. & CRABTREE, G. R. 2007. Regulation of dendritic development by neuron-specific chromatin remodeling complexes. *Neuron,* 56**,** 94-108.

WU, W. H., ALAMI, S., LUK, E., WU, C. H., SEN, S., MIZUGUCHI, G., WEI, D. & WU, C. 2005. Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. *Nat Struct Mol Biol,* 12**,** 1064-71.

XIAO, T., KAO, C. F., KROGAN, N. J., SUN, Z. W., GREENBLATT, J. F., OSLEY, M. A. & STRAHL, B. D. 2005. Histone H2B ubiquitylation is associated with elongating RNA polymerase II. *Mol Cell Biol,* 25**,** 637-51.

XIE, R., EVERETT, L. J., LIM, H. W., PATEL, N. A., SCHUG, J., KROON, E., KELLY, O. G., WANG, A., D'AMOUR, K. A., ROBINS, A. J., WON, K. J., KAESTNER, K. H. & SANDER, M. 2013. Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. *Cell Stem Cell,* 12**,** 224-37.

XUE, Y., WONG, J., MORENO, G. T., YOUNG, M. K., COTE, J. & WANG, W. 1998. NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell,* 2**,** 851-61.

YAMADA, K., FROUWS, T. D., ANGST, B., FITZGERALD, D. J., DELUCA, C., SCHIMMELE, K., SARGENT, D. F. & RICHMOND, T. J. 2011. Structure and mechanism of the chromatin remodelling factor ISW1a. *Nature,* 472**,** 448-53.

YAO, Y. L. & YANG, W. M. 2003. The metastasis-associated proteins 1 and 2 form distinct protein complexes with histone deacetylase activity. *J Biol Chem,* 278**,** 42560-8.

YEN, K., VINAYACHANDRAN, V., BATTA, K., KOERBER, R. T. & PUGH, B. F. 2012. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell,* 149**,** 1461-73.

ZANIVAN, S., KRUEGER, M. & MANN, M. 2012. In vivo quantitative proteomics: the SILAC mouse. *Methods Mol Biol,* 757**,** 435-50.

ZENG, L., YAP, K. L., IVANOV, A. V., WANG, X., MUJTABA, S., PLOTNIKOVA, O., RAUSCHER, F. J., 3RD & ZHOU, M. M. 2008. Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. *Nat Struct Mol Biol,* 15**,** 626-33.

ZENG, L., ZHANG, Q., LI, S., PLOTNIKOV, A. N., WALSH, M. J. & ZHOU, M. M. 2010. Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature,* 466**,** 258-62.

ZHANG, Y., LEROY, G., SEELIG, H. P., LANE, W. S. & REINBERG, D. 1998. The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell,* 95**,** 279-89.

ZHANG, Y., MOQTADERI, Z., RATTNER, B. P., EUSKIRCHEN, G., SNYDER, M., KADONAGA, J. T., LIU, X. S. & STRUHL, K. 2009. Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat Struct Mol Biol,* 16**,** 847-52.

ZHANG, Y., NG, H. H., ERDJUMENT-BROMAGE, H., TEMPST, P., BIRD, A. & REINBERG, D. 1999. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev,* 13**,** 1924-35.

ZHANG, Z., WIPPO, C. J., WAL, M., WARD, E., KORBER, P. & PUGH, B. F. 2011. A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science,* 332**,** 977-80.

ZHOU, W., ZHU, P., WANG, J., PASCUAL, G., OHGI, K. A., LOZACH, J., GLASS, C. K. & ROSENFELD, M. G. 2008. Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Mol Cell,* 29**,** 69-80.

ZHU, J., ADLI, M., ZOU, J. Y., VERSTAPPEN, G., COYNE, M., ZHANG, X., DURHAM, T., MIRI, M., DESHPANDE, V., DE JAGER, P. L., BENNETT, D. A., HOUMARD, J. A., MUOIO, D. M., ONDER, T. T., CAMAHORT, R., COWAN, C. A., MEISSNER, A., EPSTEIN, C. B., SHORESH, N. & BERNSTEIN, B. E. 2013. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell,* 152**,** 642-54.

ZHU, P., ZHOU, W., WANG, J., PUC, J., OHGI, K. A., ERDJUMENT-BROMAGE, H., TEMPST, P., GLASS, C. K. & ROSENFELD, M. G. 2007. A histone H2A deubiquitinase complex coordinating histone acetylation and H1 dissociation in transcriptional regulation. *Mol Cell,* 27**,** 609-21.

ZHUANG, T., HESS, R. A., KOLLA, V., HIGASHI, M., RAABE, T. D. & BRODEUR, G. M. 2014. CHD5 is required for spermiogenesis and chromatin condensation. *Mech Dev,* 131**,** 35-46.

ZIV, Y., BIELOPOLSKI, D., GALANTY, Y., LUKAS, C., TAYA, Y., SCHULTZ, D. C., LUKAS, J., BEKKER-JENSEN, S., BARTEK, J. & SHILOH, Y. 2006. Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. *Nat Cell Biol,* 8**,** 870-6.